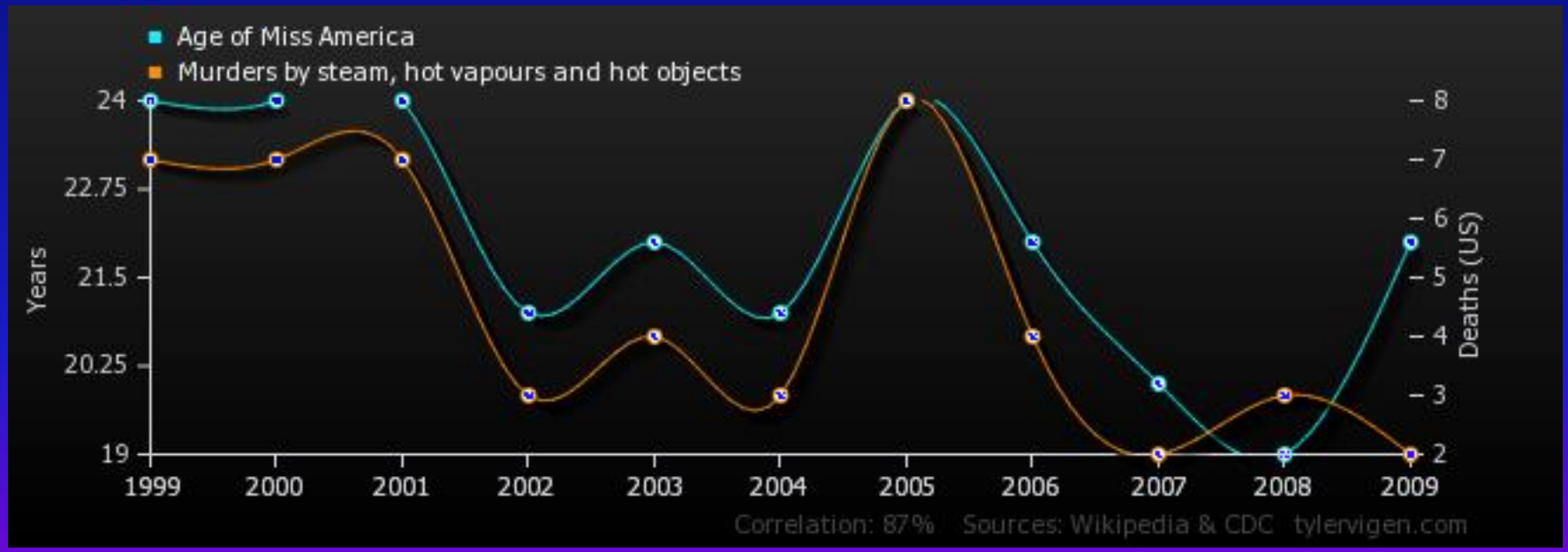
Correlations: Assessing Shared Variance



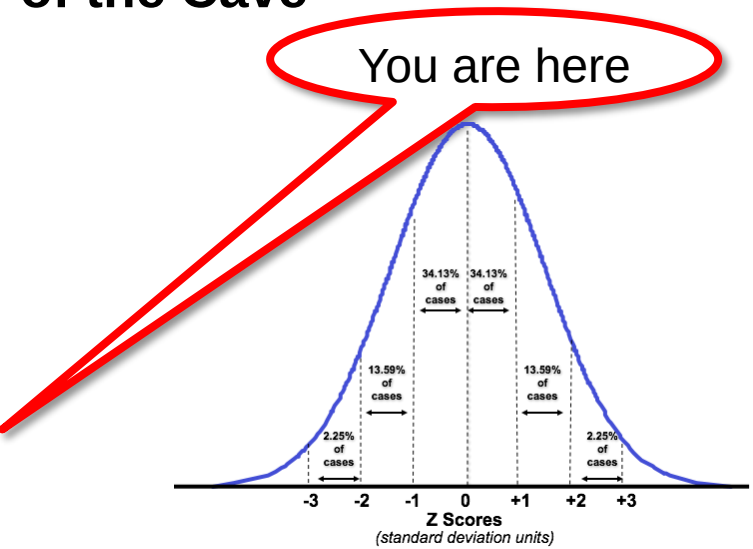
Correlations: http://tylervigen.com/view_correlation?id=2948assessing shared variance

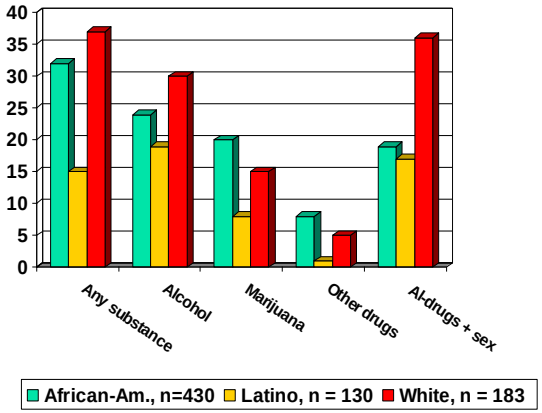
Click the image for more…

Dr. David McKirnan, davidmck@uic.edu

SHOUT
SHARE
IT

1. Introduction to statistics & number scales
2. The Z score and the normal distribution
3. The logic of research; Plato's Allegory of the Cave
4. Testing hypotheses: The critical ratio
5. Calculating a t score
6. Testing t: The Central Limit Theorem
7. Correlations: Measures of association





© Dr. David J. McKirnan, 2014
The University of Illinois Chicago
McKirnanUIC@gmail.com
Do not use or reproduce without permission

Testing a hypothesis: t-test versus correlation

How much do you love statistics?

A = Completely

B = A lot

C = Pretty much

D = Just a little

E = Not at all

Dr. David McKirnan, davidmck@uic.edu

How are you doing in statistics?

A = Terrific

B = Good

C = OK

D = Getting by

E = Not so good

How could we test the hypothesis that statistical love affects actual performance…

Shutter stock

- Using an experimental design?
- Using a correlational or measurement design?

Testing a hypothesis: t-test versus correlation

- **t-tests:**
  - **Used for experiments**
  - **Manipulate the independent variable**
    - Sexy versus "placebo" statistics instructors
  - **Mesasure differences in the Dependent Variable.**
    - Statistics module grade.
- **Correlations:**
  - **Used for measurement studies**
  - **Measure the Predictor variable**
    - How much do people love statistics
  - **Measure the Outcome variable**
    - Statistics grade.

An experimental approach

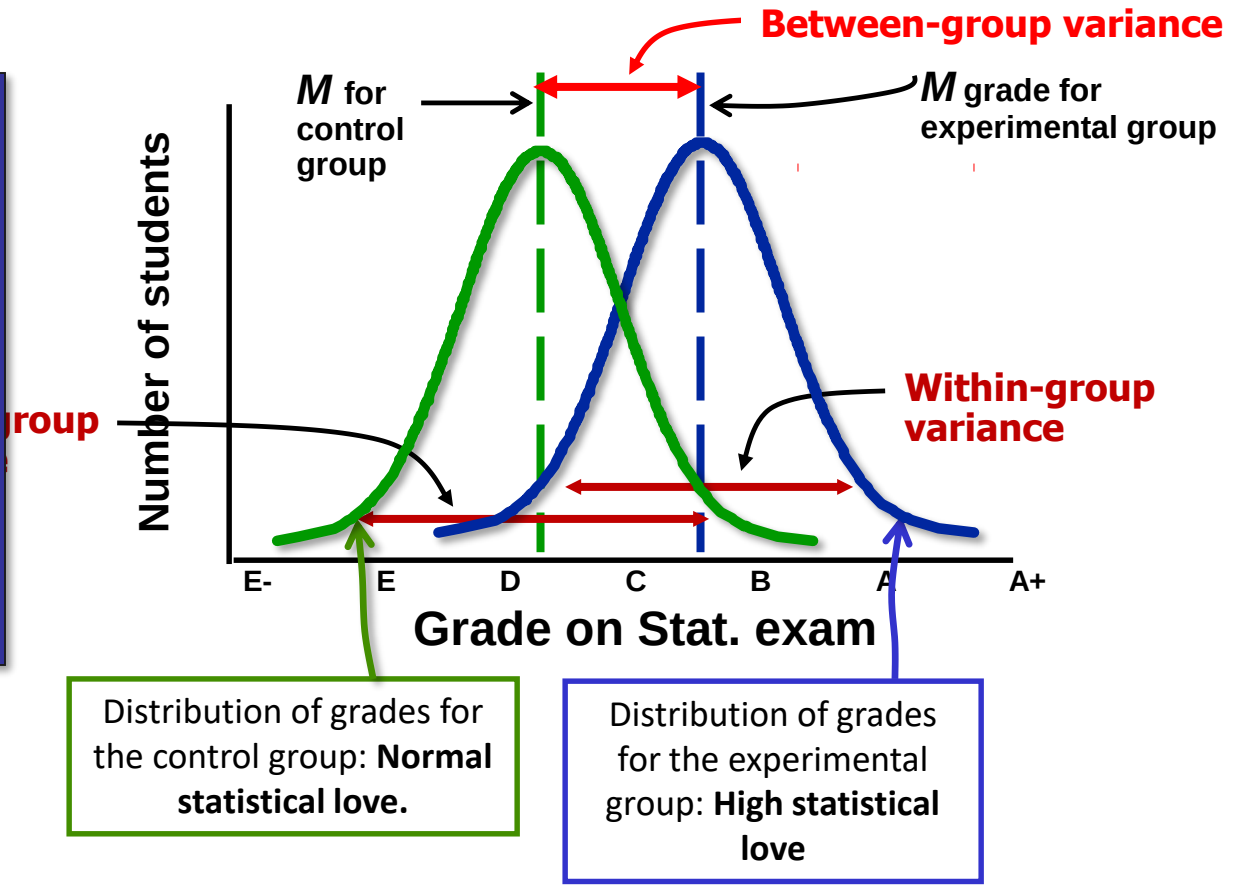
Hypothesis: Using sexy statistics instructors to induce love of statistics will lead to higher grades

Statistical question: Did the experimental group get statistically significantly higher grades than did the control group?

t-test:

How much variance is there **between** the **groups**

What would we expect by ***chance*** given the amount of variance **within groups**.



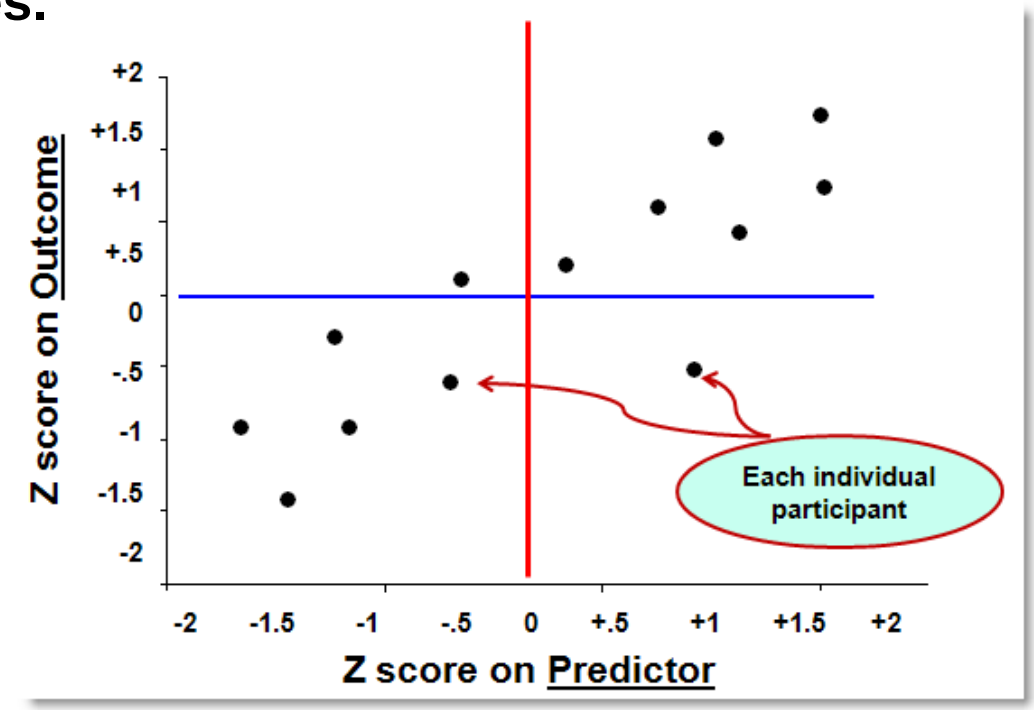
Dr. David McKirnan, davidmck@uic.edu

Taking a correlation approach

Statistics students like their statistics instructor.

Correlation:

- ✓ **Are naturally occurring Individual differences on the Predictor Variable.**
- ✓ **Associated with individual differences on the Outcome?**
- ✓ **To test this we examine each participants' scores on the two** (measured) **study variables.**

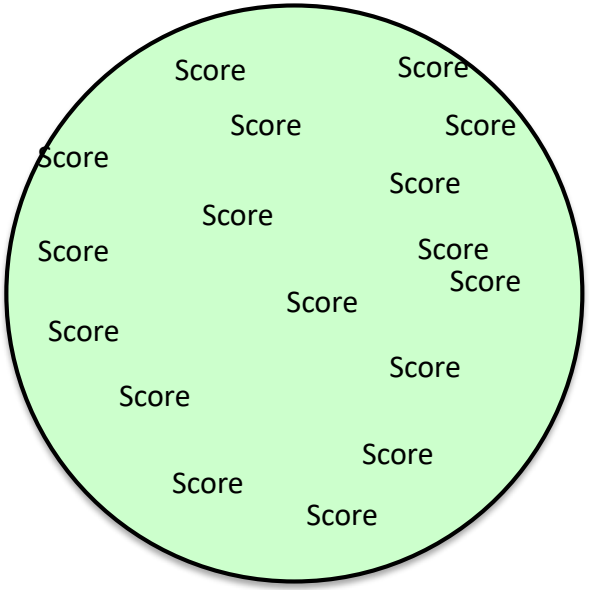


Dr. David McKirnan, davidmck@uic.edu

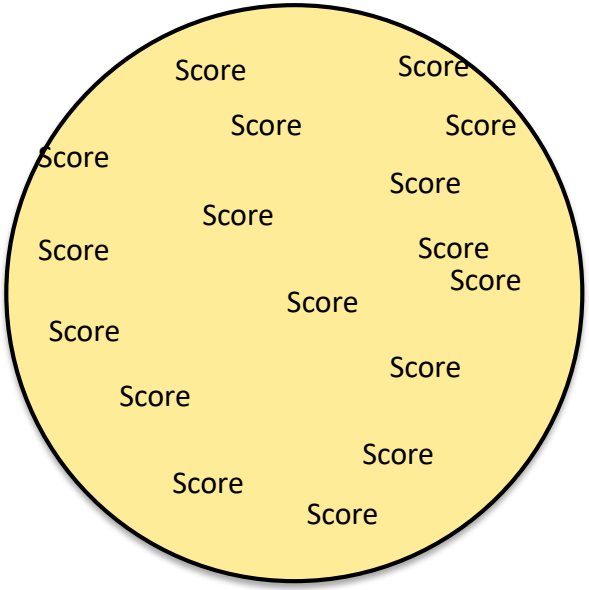
Venn diagrams: The logic of the correlation

Imagine we take all the scores of the class on a measure of Statistical Love.

We can array them showing how much variance there is among students…



We can illustrate the variance by imagining a circle that contains all the scores

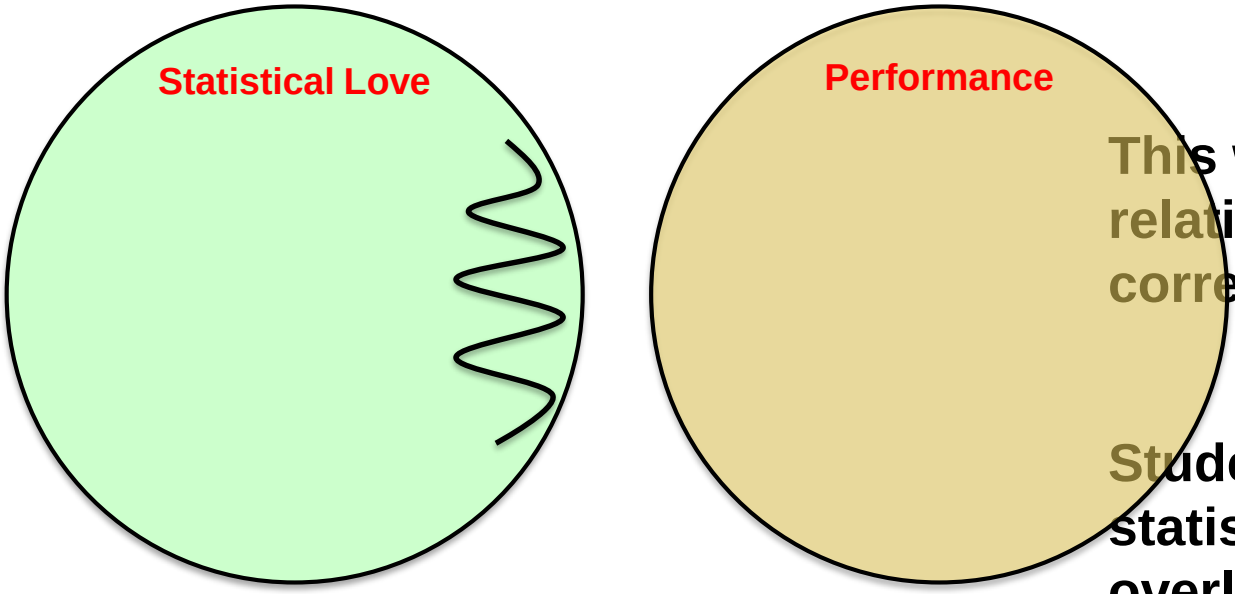


We could do the same thing with our performance scores.

Venn diagrams: The logic of the correlation

A correlation tests how much these scores overlap, i.e., their Shared Variance.

As they share more variance – i.e. students who are high on one score are also high on the other and visa-versa – the circles overlap.



This would illustrate a relatively low correlation.

Students scores on statistical love do not overlap much with their scores on performance.

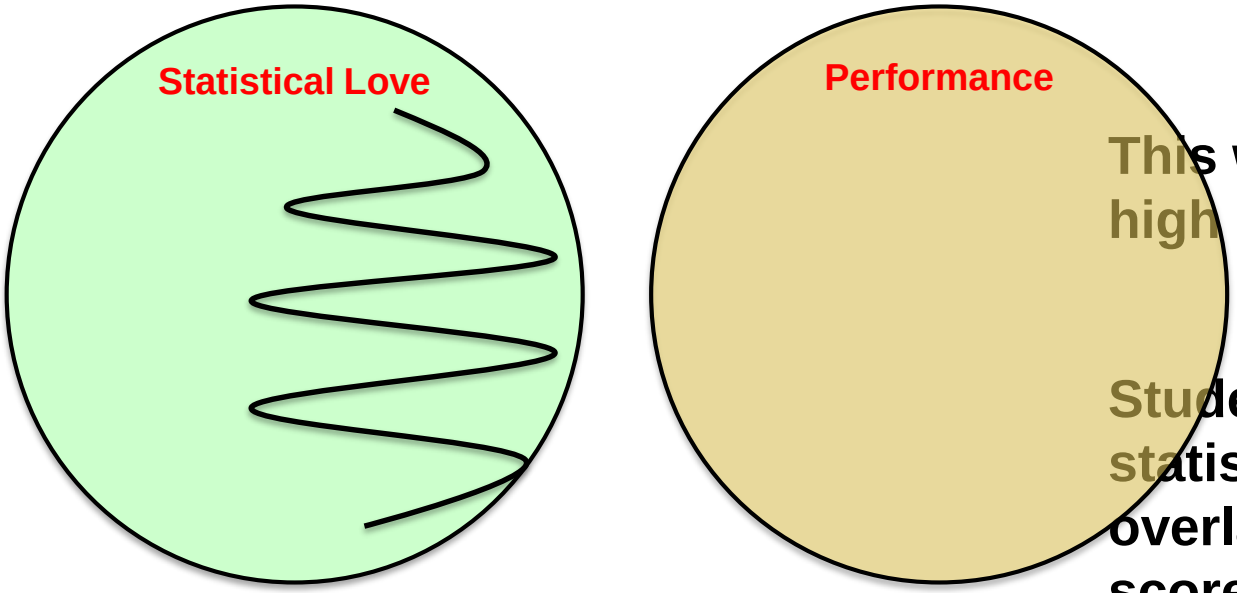
Dr. David McKirnan, davidmck@uic.edu

Venn diagrams: The logic of the correlation

If students who were:

- **high on statistical love were equally high on performance**
- **or low on love equally low on the performance**

The two measures would share a lot of variance.



This would illustrate a high correlation.

Students scores on statistical love overlap a lot with their scores on performance.

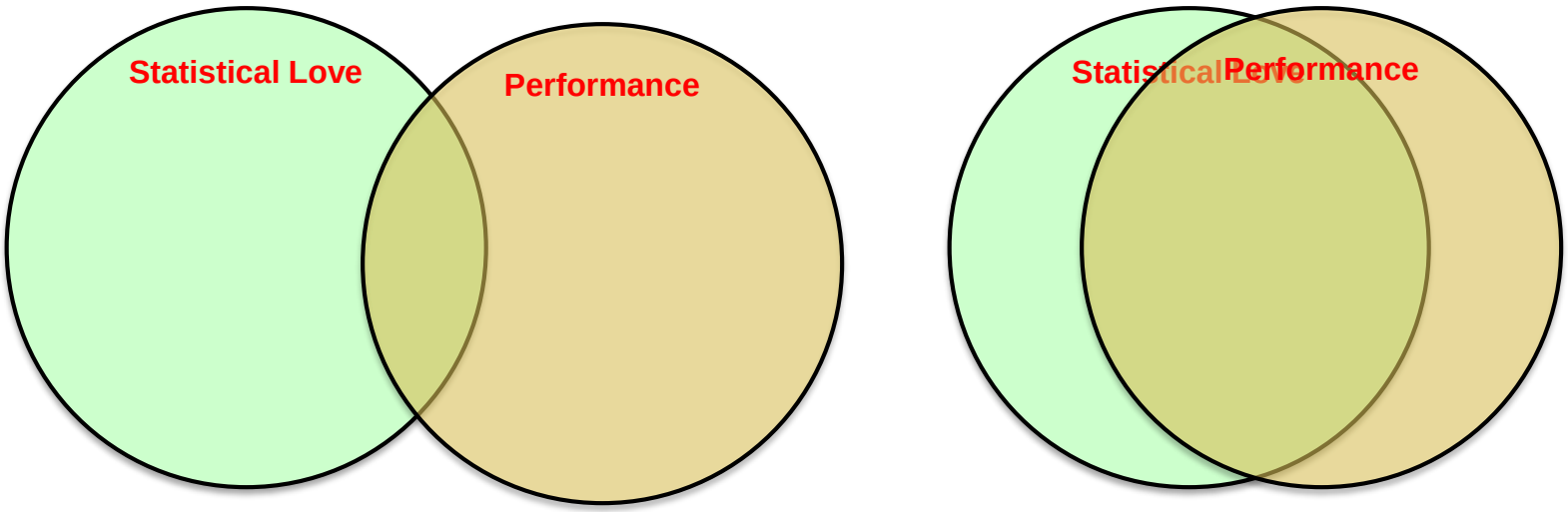
Dr. David McKirnan, davidmck@uic.edu

Venn diagrams: The logic of the correlation

We test how much two variables share their variance by computing participants' Z score on each measure.

> **Z = how much a participant is above or below the M, divided by the Standard Deviation.**

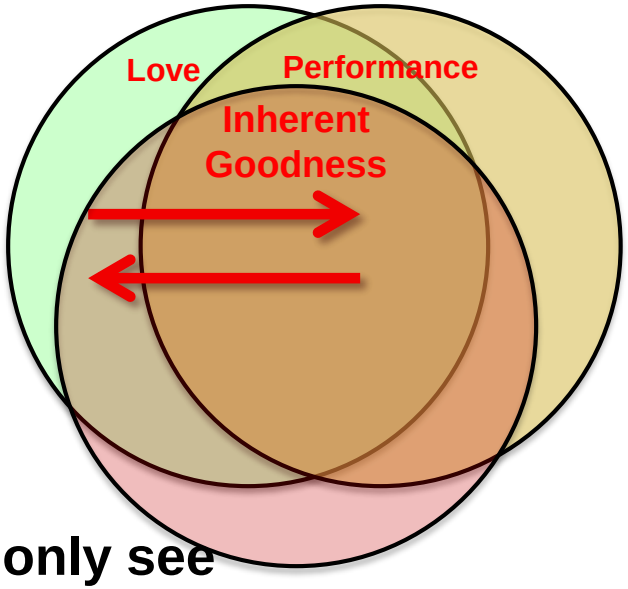
Comparing participants' Zs allows us to derive the correlation coefficient.



Dr. David McKirnan, davidmck@uic.edu

Venn diagrams: The logic of the correlation

Let's say our results look like this:



Can we conclude that love of statistics causes higher performance?

Of course the causal arrow may go the other way…

Or, both ways.

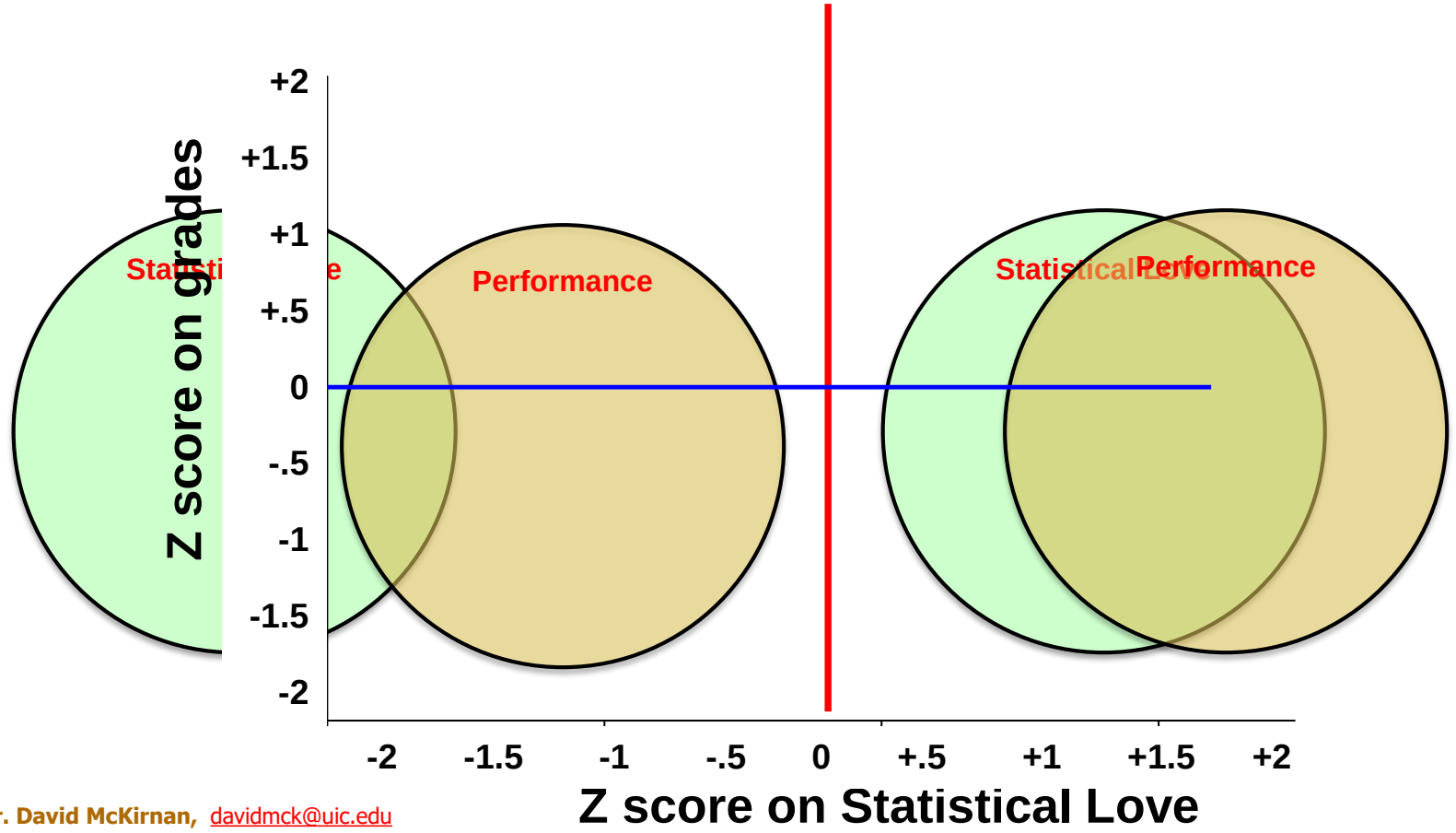
- **When we collect cross- sectional data we only see a 'snapshot' of attitudes or behavior.**
- **Longitudinal data may show us that over time**
  - **Love leads to performance,**
  - **Performance → more love, etc.**

Finally, a third variable – such as inherent goodness as a person – may cause both statistical love and performance.

Dr. David McKirnan, davidmck@uic.edu

Conceptually, the correlation asks how much the variance on two measures is overlapping – or shared –within a set of participants.

We test statistically this by displaying each persons scores on the two variables in a Scatter Plot, and deriving Z scores for them.



Dr. David McKirnan, davidmck@uic.edu

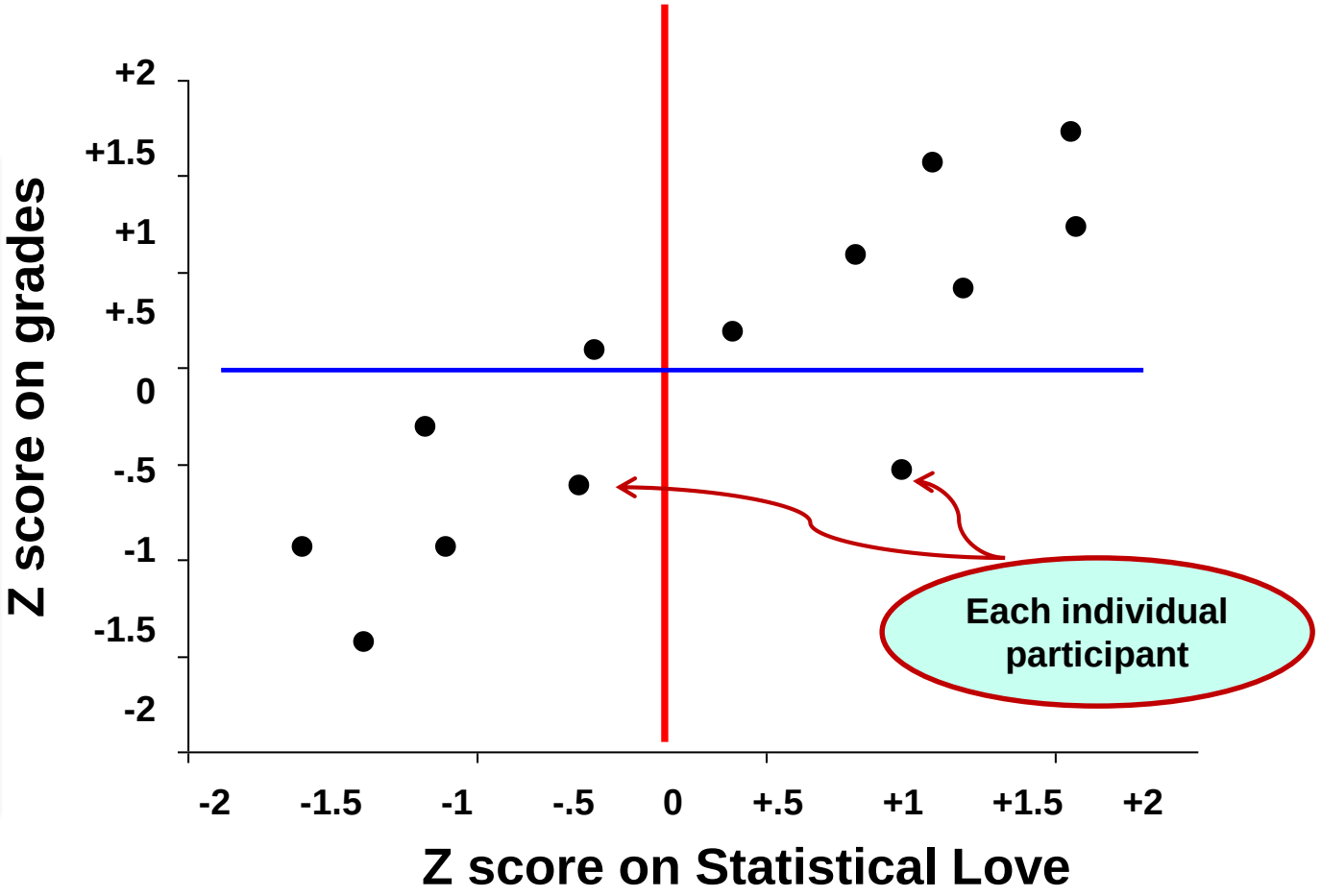
Hypothesis: Students who are "naturally" high in Love of Statistics will have higher grades

Statistical question: Are students who are higher on a measure of Statistical Love also higher on the outcome measure (grades)?

Correlation:

How much is **variance** on the predictor variable **shared** with the outcome variable:

Are students who are high in Stat Love high by a similar amount on grades?



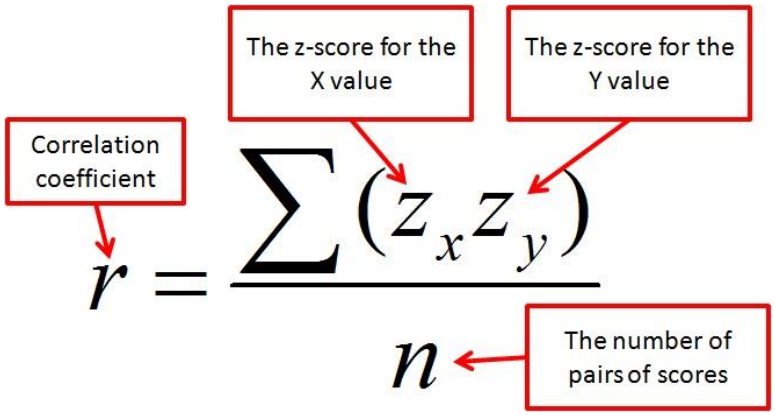
Dr. David McKirnan, davidmck@uic.edu

Pearson Correlation Coefficient:

…measures how similar the variance is between two variables, a.k.a. "shared variance".

- Are people who are above or below the mean on one variable similarly above or below the M on the second variable?
  - If everyone who is a certain amount over the M on Statistical Love (say, Z ≈ +1.5)…
  - …is about the same amount above the M on performance (Z also ≈+1.5)…
  - …the correlation would be +1.0.

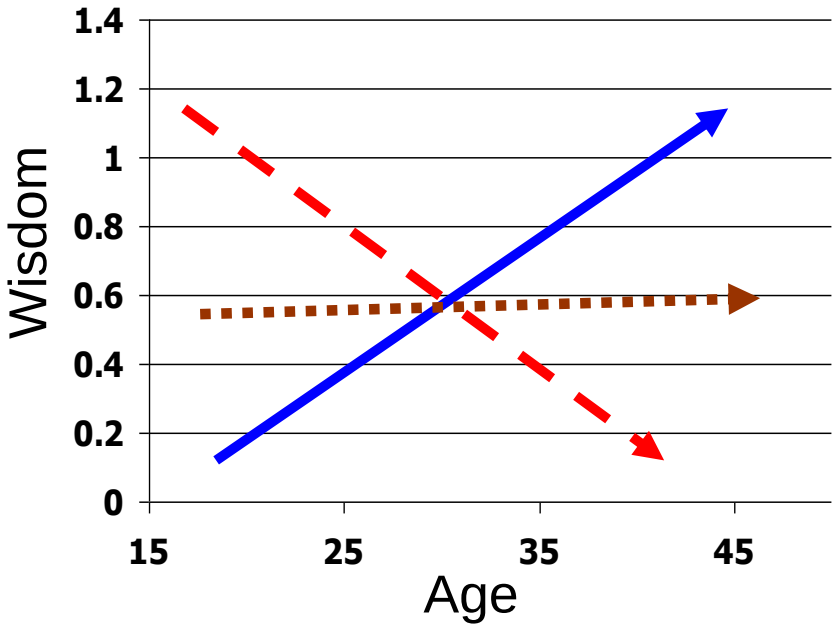
We assess shared variance by multiplying the person's Z scores for each of the two variables / n :


The z-score for the X value
The z-score for the Y value
Correlation coefficient
$r = \frac{\sum (z_x z_y)}{n}$
The number of pairs of scores

The Pearson Correlation coefficient:

Measures linear relation of one variable to another within participants, e.g. *Wisdom* & *Age*;

- **Cross-sectional: how much are participants' wisdom scores related to their different ages?**
- **Longitudinal: how much does wisdom increase (or decrease) as people age?**



Positive correlation: among older participants wisdom is higher…

Negative correlation: older participants actually show *lower* wisdom…

No (or low) correlation: higher / lower age makes no difference for wisdom..

Let's imagine we plot each participant's wisdom score, on a scale of '0' to '1.4', against his / her age (ranges from 15 to 50)

These variables have different scales ['0' → '1.4' versus 15 → 50].

How can we make them comparable?

We can standardize the scores by turning them into Z scores.

- ✓ **Calculate the M and S for each variable**
- ✓ **Express each person's score as their Z score on that variable.**

Imagine the M age = 30, Standard Deviation [S] = 10

What age would be one standard deviation above the Mean (Z = 1)?

A Z = 20

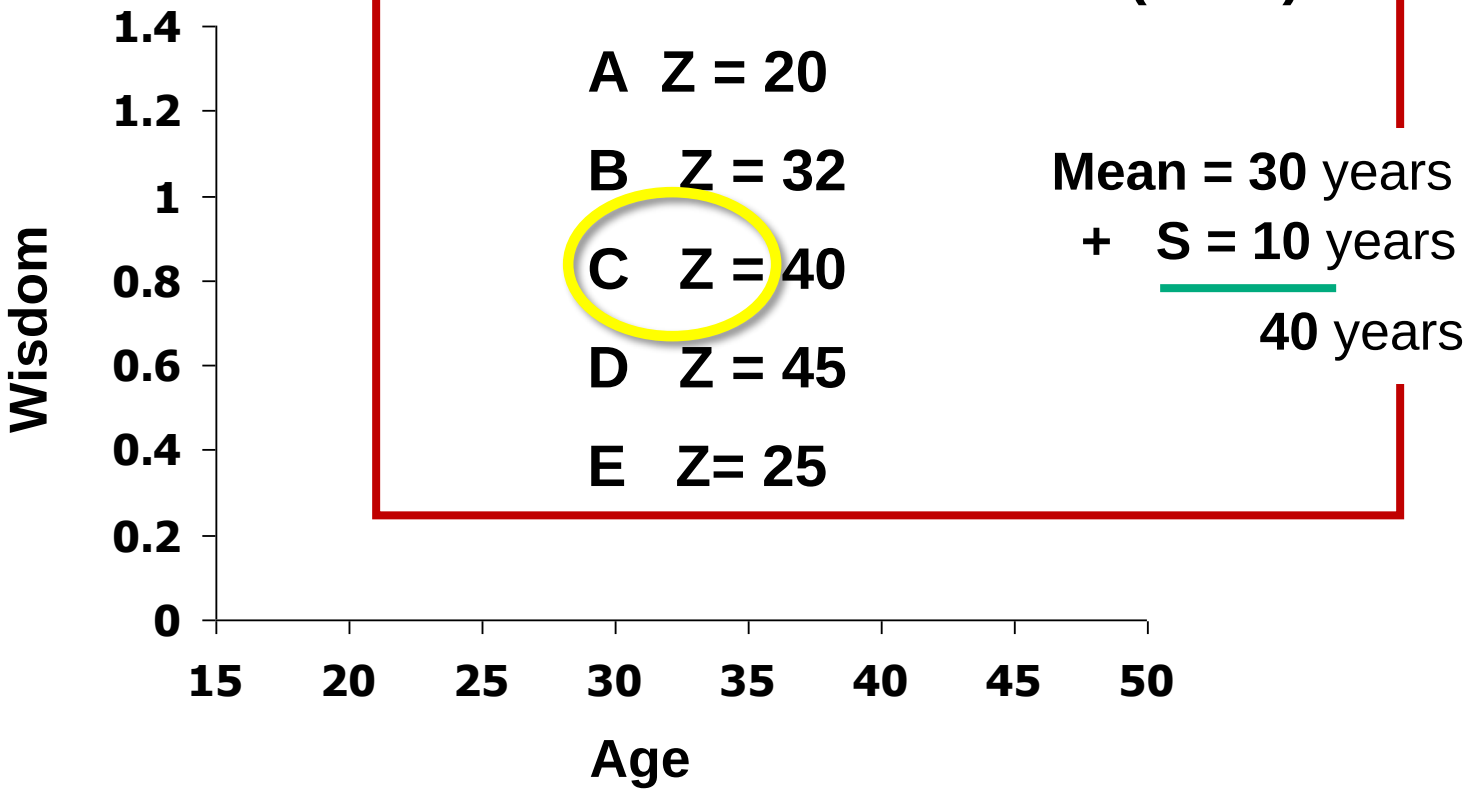
B Z = 32

C Z = 40

D Z = 45

E Z= 25

Mean = 30 years
\+ **S = 10** years
40 years



Dr. David McKirnan, davidmck@uic.edu

Imagine the M age = 30, Standard Deviation [S] = 10

Participant 14 is age 25. What is her Z score on age?

A = 0

B = +.5

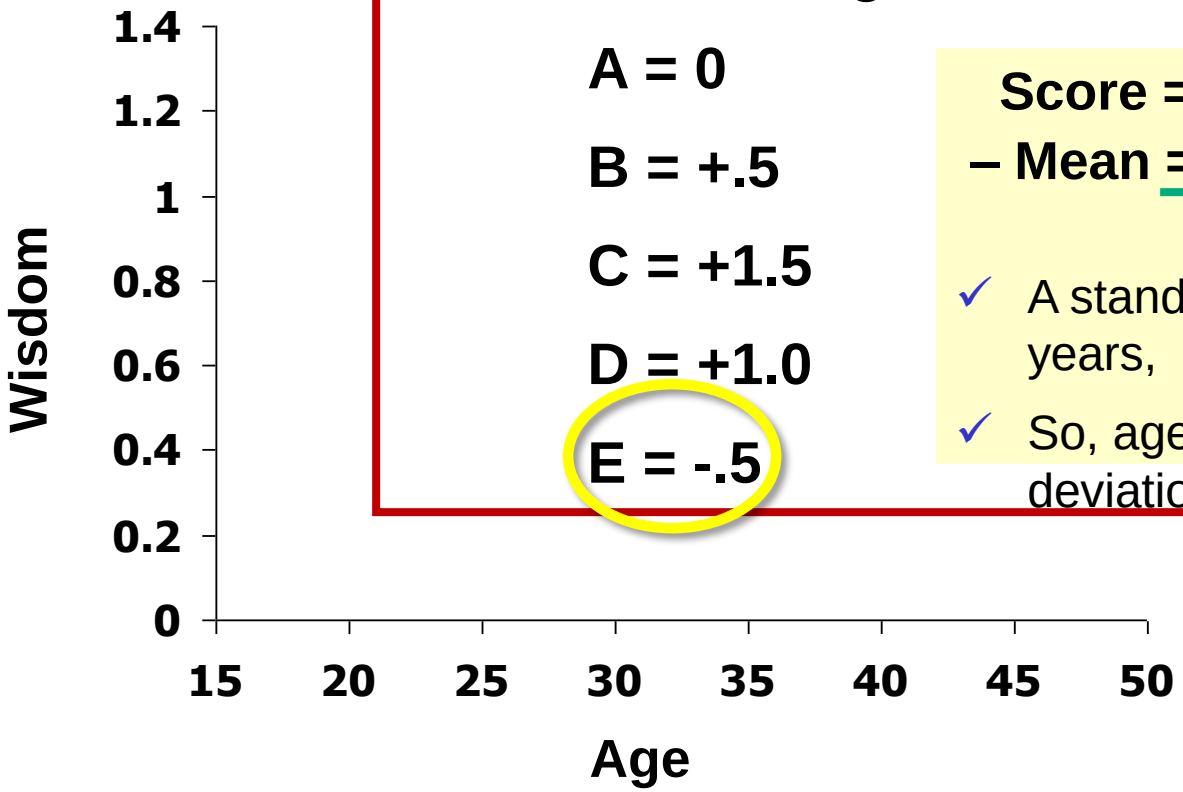
C = +1.5

D = +1.0

E = -.5

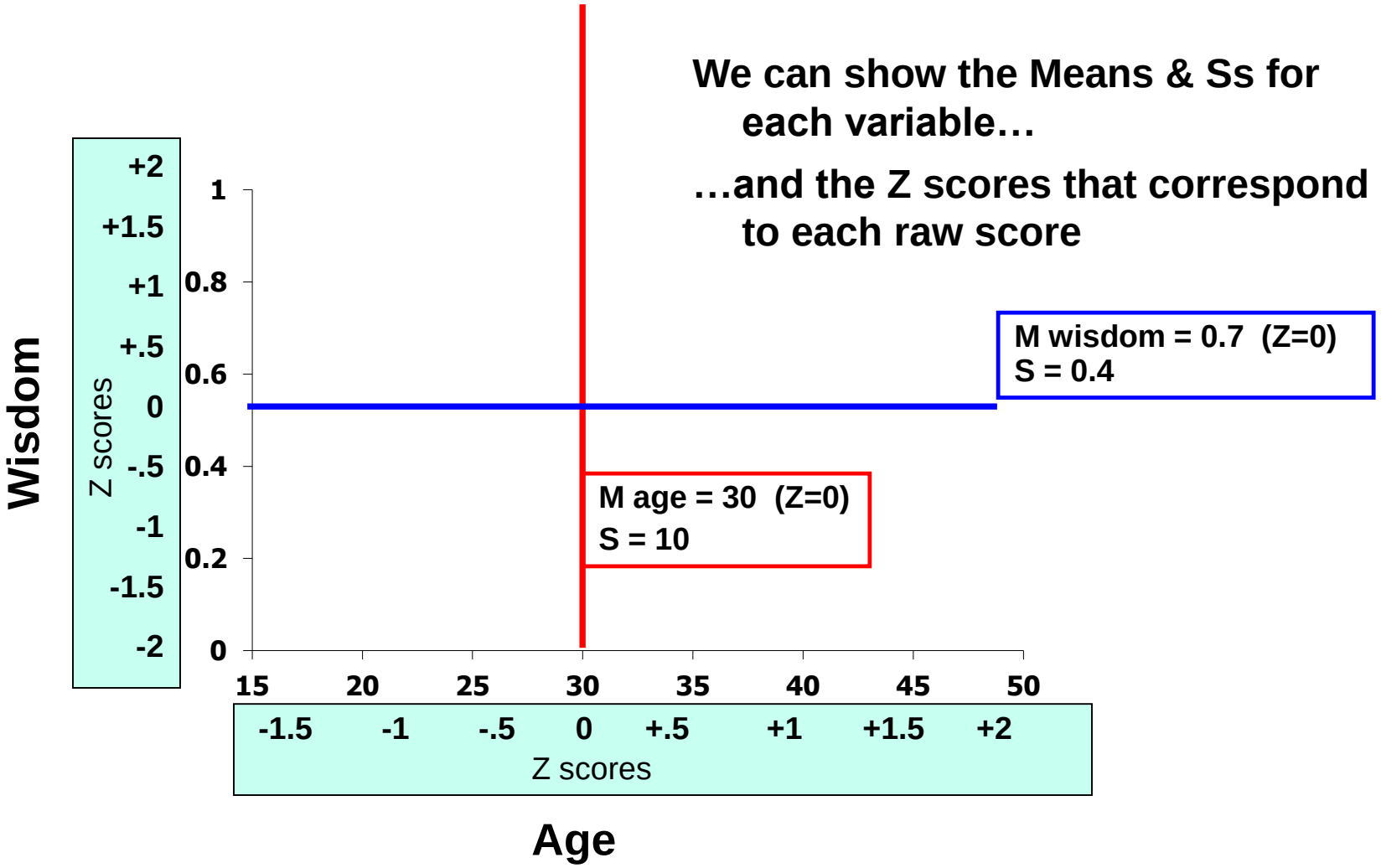
Score = 25 years
– Mean = 30 years
5 years below the mean

- ✓ A standard deviation =10 years,
- ✓ So, age 25 is ½ standard deviation below the mean.

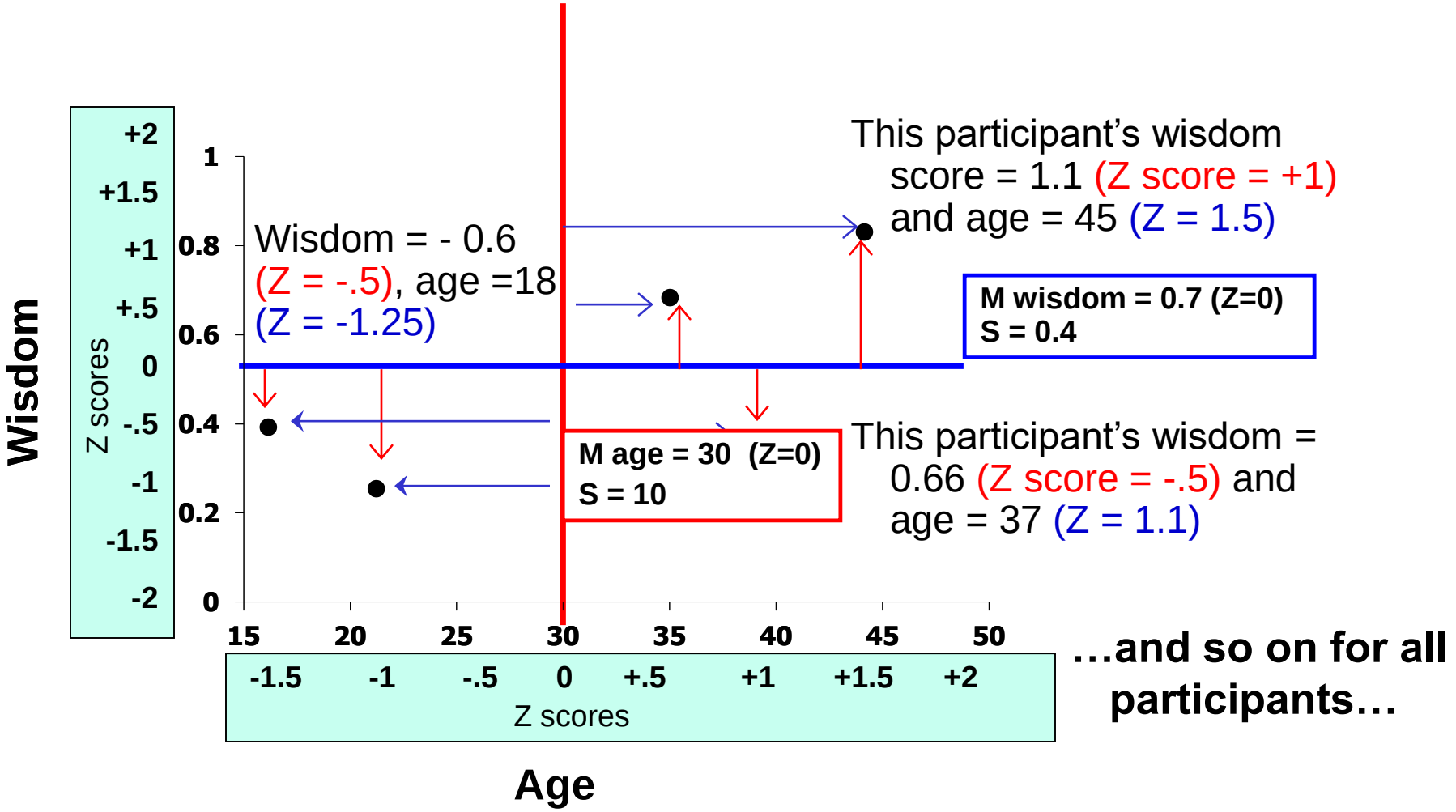


Age: **Mean [M] = 30, Standard Deviation [S] = 10**

Wisdom: **M = 0.7, S = 0.4**

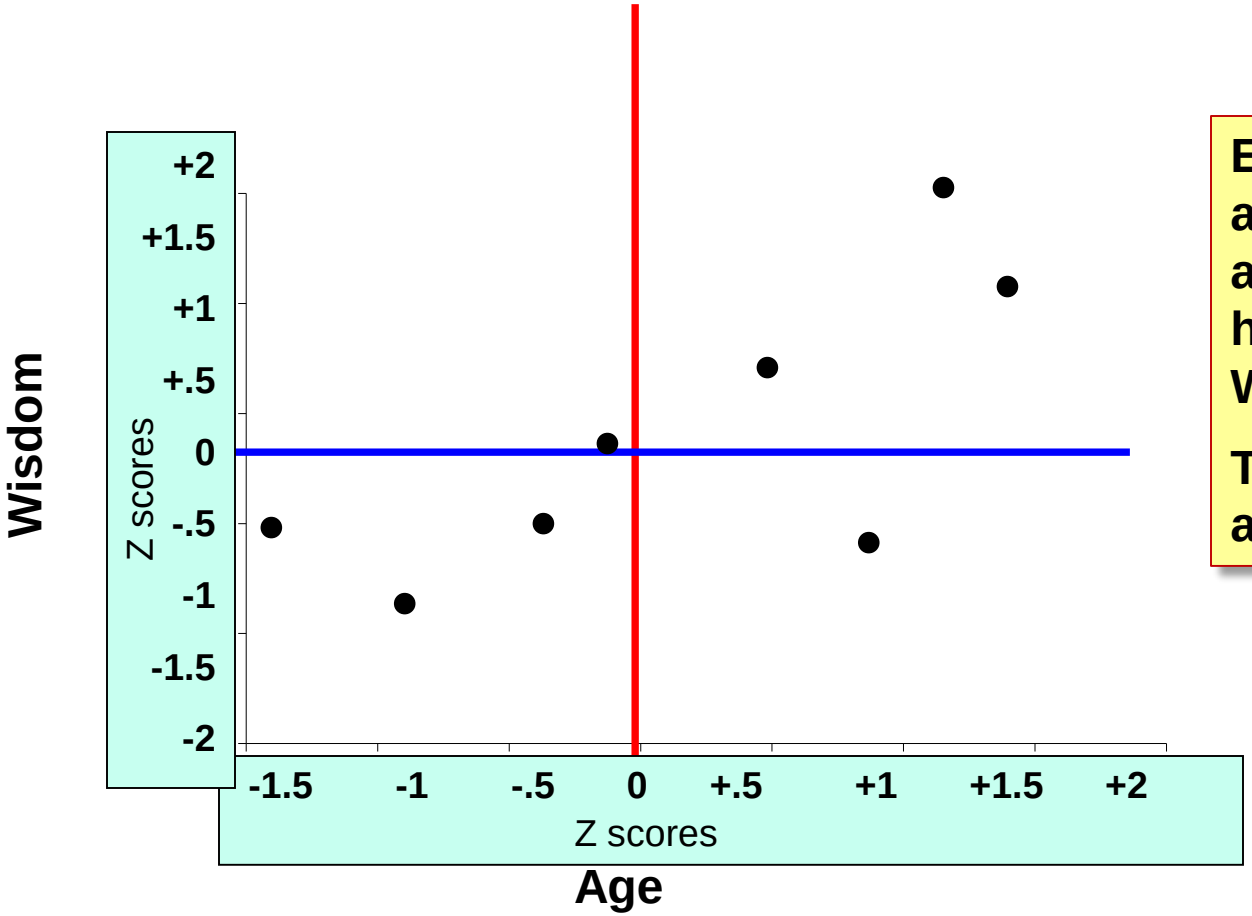


- **For each participant we use Z scores to show how far above or below the Mean they are on the two variables.**
- **Z scores allow us to standardize and compare the variables.**



…and so on for all participants…

- For each participant we use Z scores to show how far above or below the Mean they are on the two variables.
- The Z scores allow us to standardize the variables.



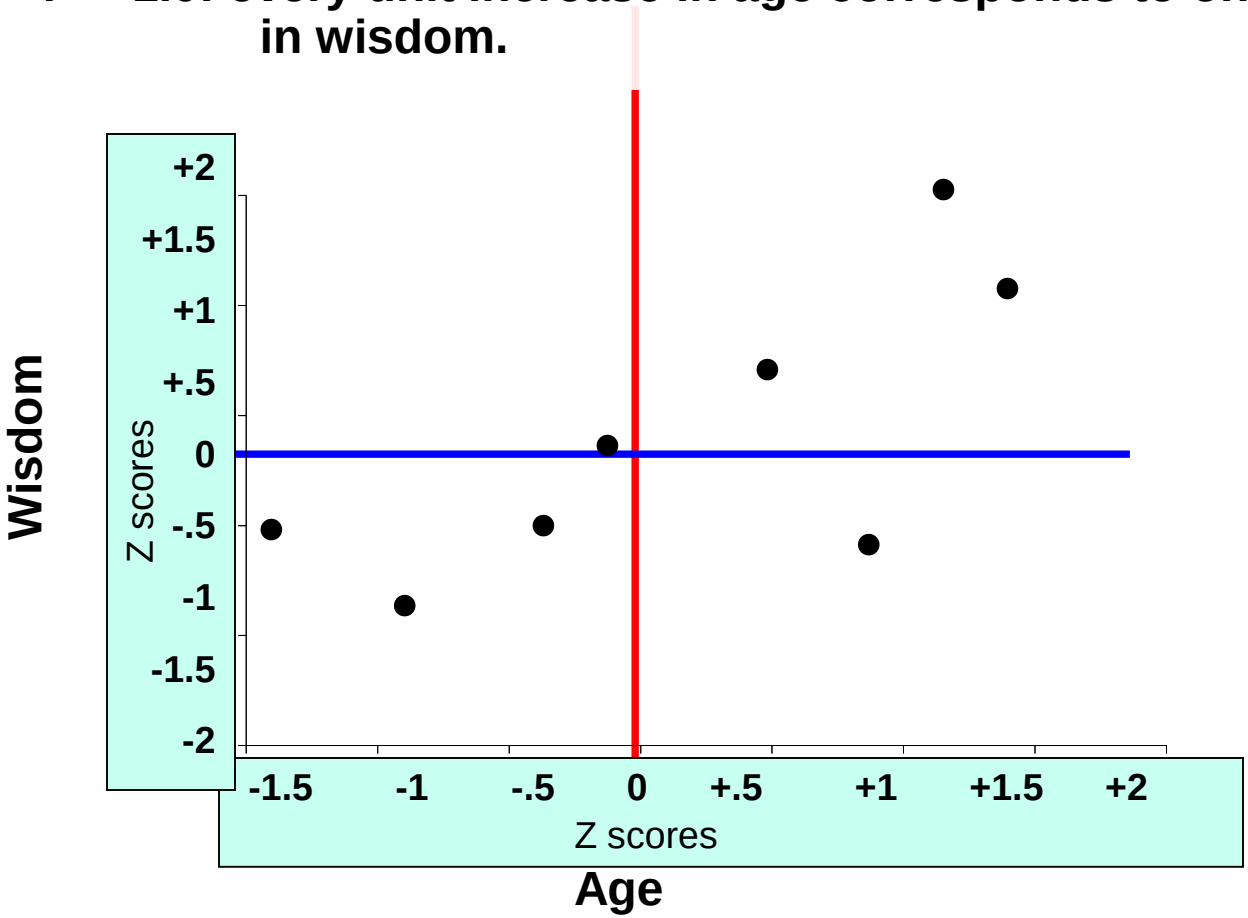
Eventually we can see a pattern of scores – as age scores get higher so does Wisdom.

This would represent a positive correlation

Correlations (r) range from +1.0 to -1.0:

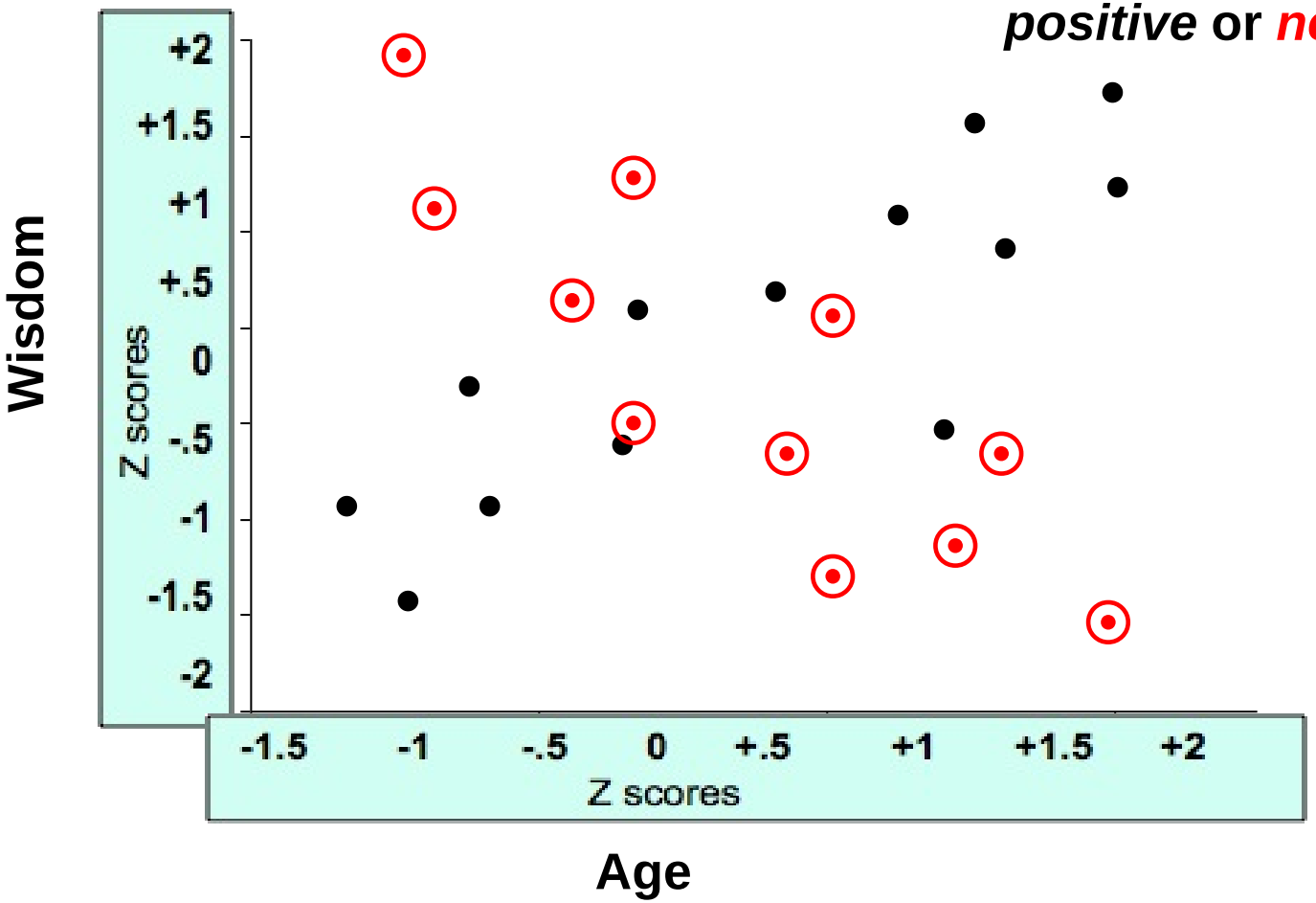
r = +1.0: for every unit increase in age there is exactly one unit increase in wisdom.

r = -1.0: every unit increase in age corresponds to one unit *decrease* in wisdom.



The pattern of Z scores for the "Y" (Wisdom) **and "X"** (age) **variables determines how strong the correlation is…**

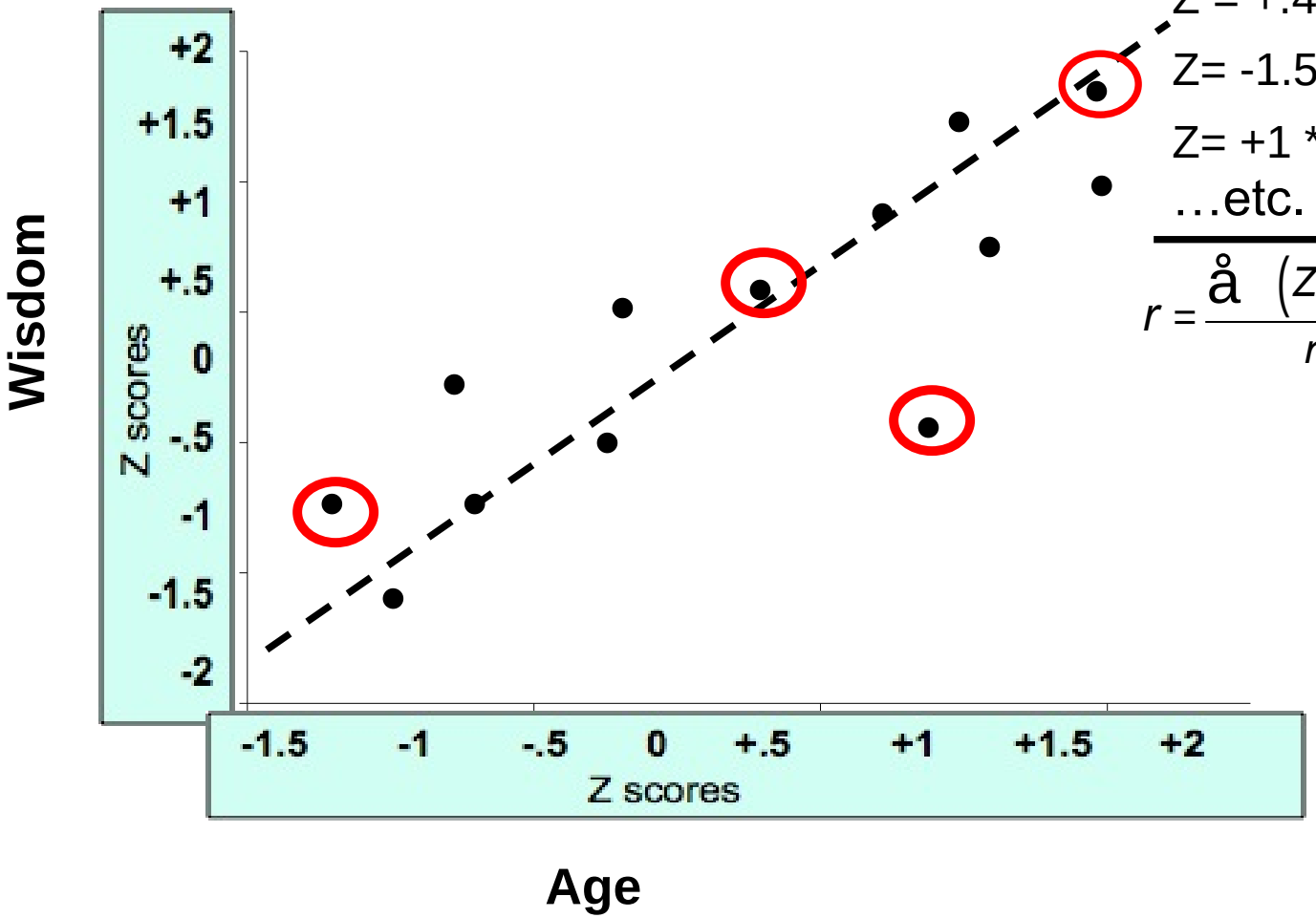
And whether it is *positive* or *negative*…



Dr. David McKirnan, davidmck@uic.edu

Correlations; larger patterns of Zs

Two variables are positively correlated if each score above / below the M on the first variable is about the same amount above / below M on Variable 2.



Z=+1.7 * Z=+1.5 = +2.55

Z = +.4 * Z =+.5 = +.2

Z= -1.5 * Z= -1 = +1.5

Z= +1 * Z= -.5 = -.5

…etc.

$$r = \frac{\text{å}\ (Zx * Zy)}{n} = .8$$

Dr. David McKirnan, davidmck@uic.edu

Correlations; larger patterns of Zs

The same logic applies for a negative correlation:

- ✓ each score above the M on age is below the M on Wisdom, and visa versa.

Z= +1.9 * Z= -1 = -1.9

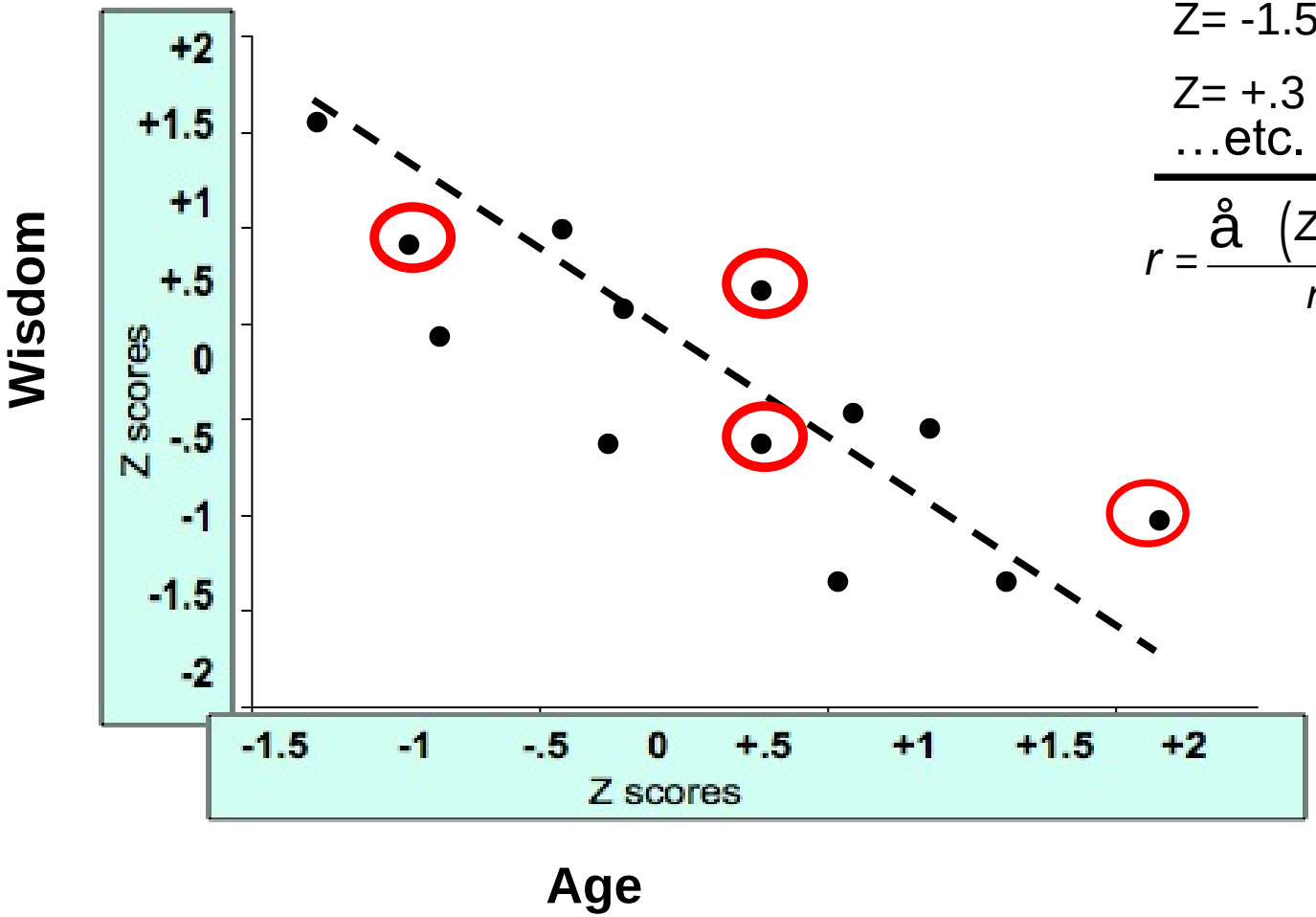
Z = +.4 * Z =+.5 = +.2

Z= -1.5 * Z= +.7 = -1.05

Z= +.3 * Z= -.5 = -.15

…etc.

$$r = \frac{\sum (Zx * Zy)}{n} = -.8$$



Dr. David McKirnan, davidmck@uic.edu

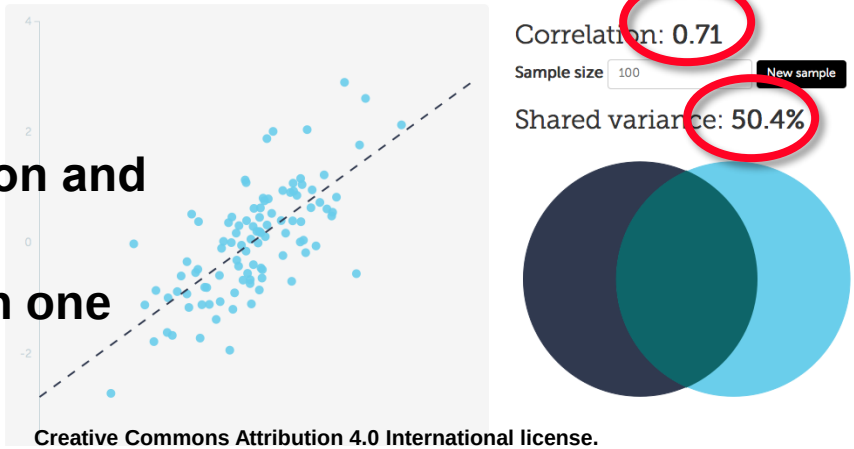
Data patterns and correlations

When two variables are strongly related

– for each person Zy ≈ Zx –

the Scatter Plot shows a "tight" distribution and r (the correlation coefficient) **is high…**

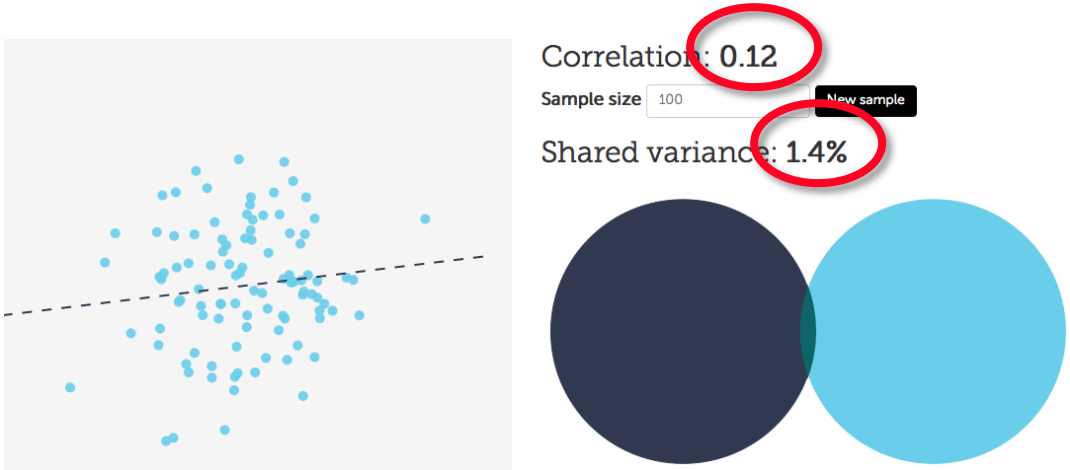
Here a substantial amount of variance in one variable is "shared" with the other.



Creative Commons Attribution 4.0 International license.

As the variables are less related, the plot gets more random or "scattered".

r – and the estimate of shared variance – go down.



Watch how correlations work!

Click either image for an interactive scatter plot program from rpsychologist.com

Dr. David McKirnan, davidmck@uic.edu

Correlation Example 1

- **Research question: what are ψ consequences of loving statistics?**
- **Hypothesis: statistical love leads to less social isolation among students.**
- **Method: *Measurement study* rather than *experiment***
  - **Measure each students' scores on 2 variables:**
    - ✓ **Love of Statistics Index**
    - ✓ **Social Isolation Scale**
  - **Test whether the two variables are significantly related; if a student is high on one, is s/he also high on the other?**
- **Data: Scores on 2 measured variables for 7 students**

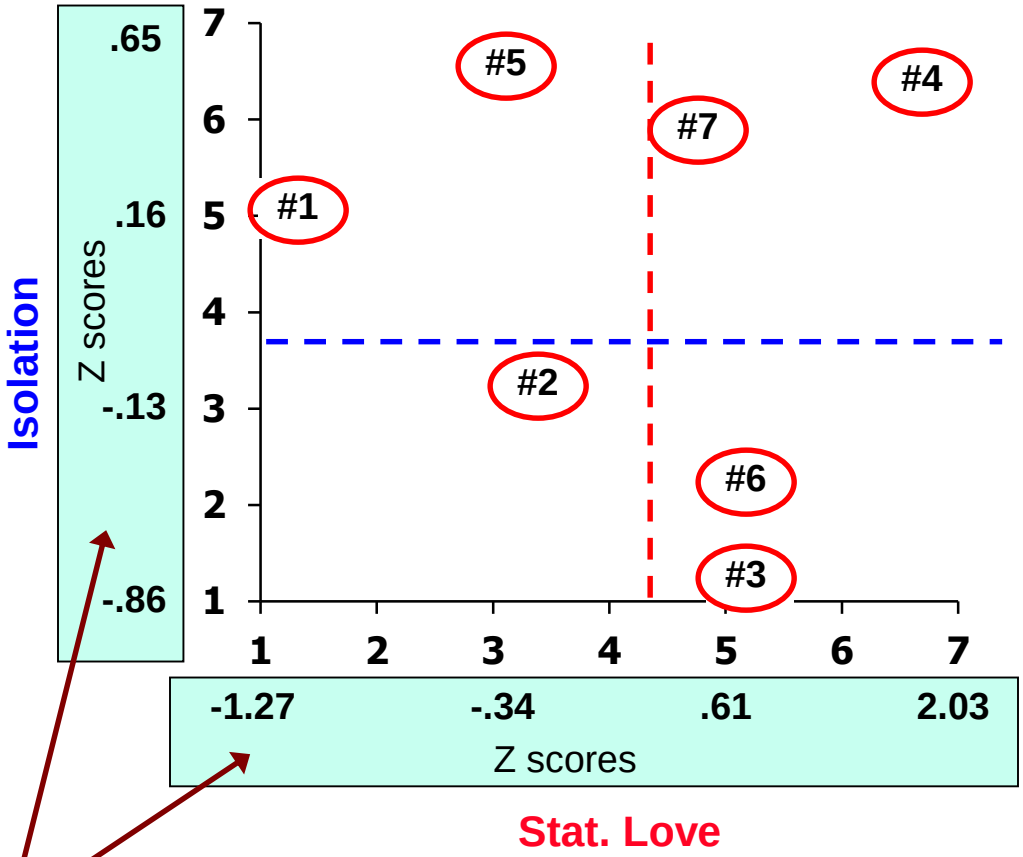
- **Imagine a data set of scores on each of the two variables for 7 people.**

Example data set 1; Raw data		
Participant	Stat. Love ('X')	Social isolation ('Y')
1	1	5
2	2	3
3	4	1
4	7	6
5	3	6
6	5	2
7	4	7
	M = 3.71	**M = 4.29**
	S = 2.12	**S = 4.49**

- **To compute a correlation we first turn these raw values into Z scores.**
- **We calculate M (Mean) and S (Standard Deviation) for each variable.**
- **Then for each participant we subtract Score – M / S to compute Z.**

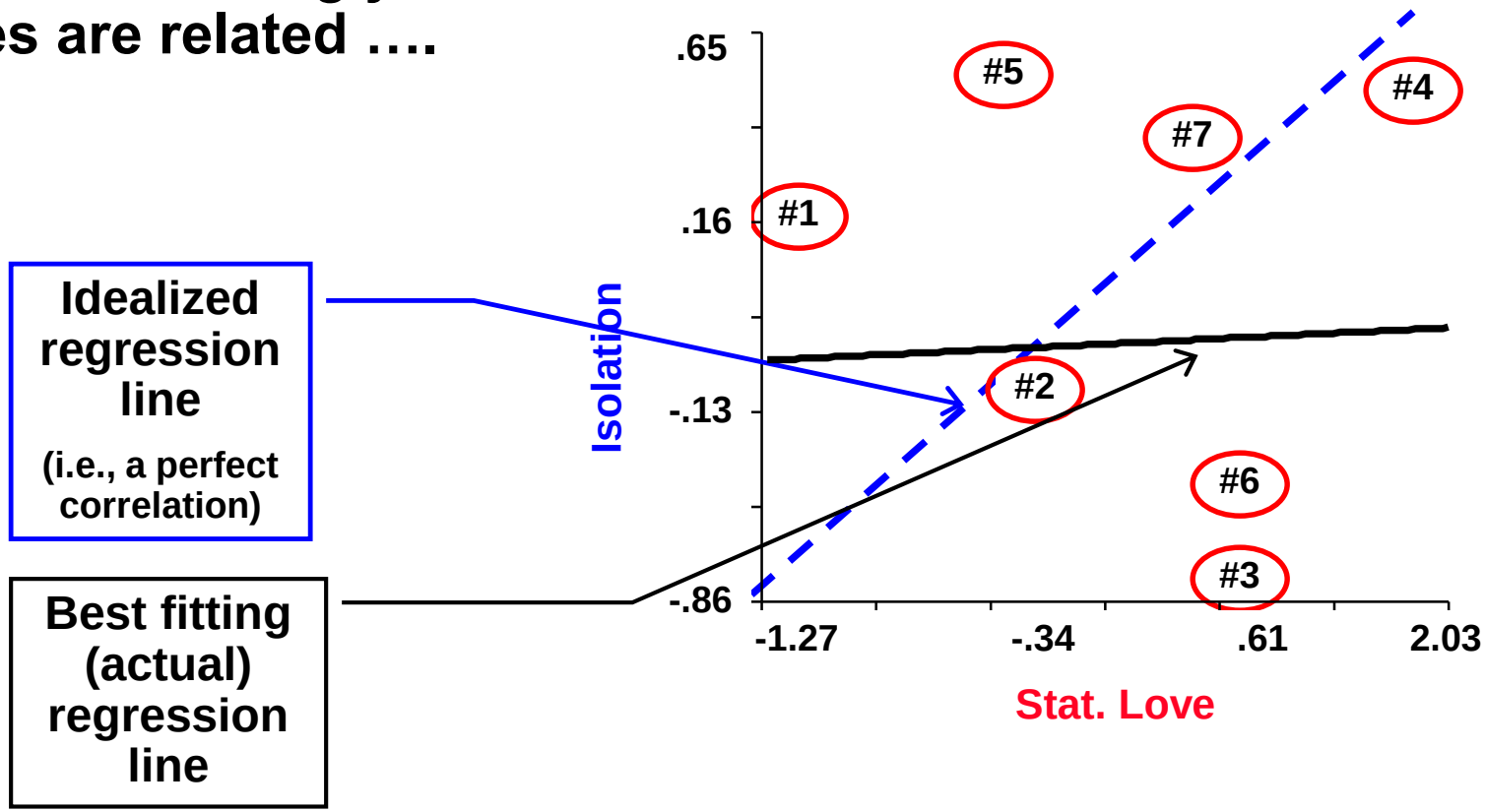
- We compute the Z scores.
- We then show the data as a Scatter Plot of individual scores.

Example data set 1; Raw data and Z scores		
Participant	Stat. Love ('X')	Isolation ('Y')
1	-1.27	.56
2	-.34	-.13
3	.14	-.86
4	2.03	.65
5	-.34	.65
6	.61	-.21
7	.34	.60



Note the different ranges of the Z scores.
This reflects different variances in the two samples.

- **The Scatter plot graphically displays how strongly the variables are related ….**

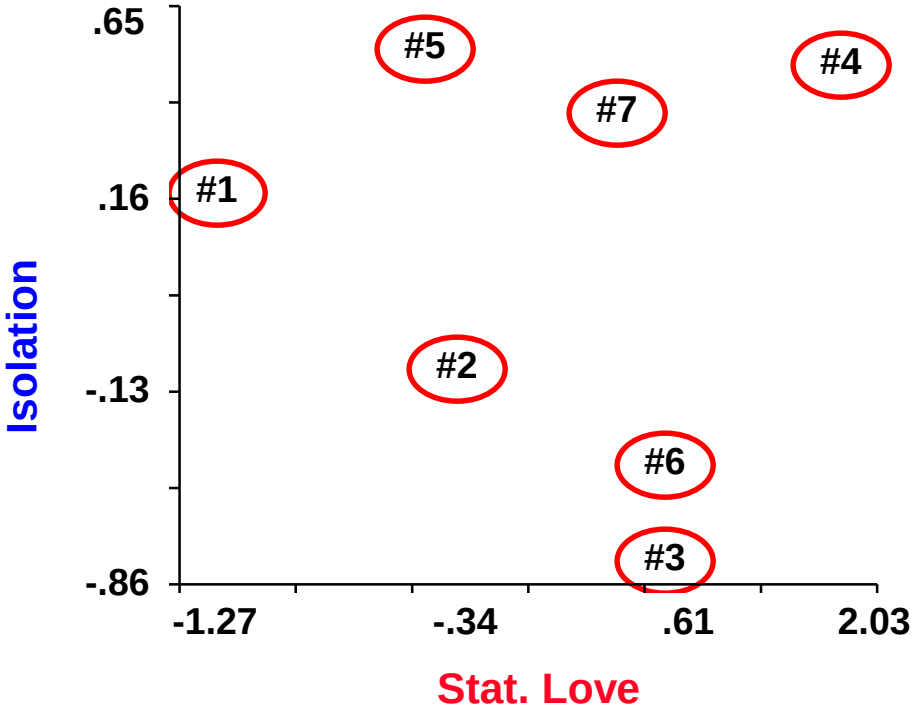


Correlation example 1; data & plot

The low correlation coefficient confirms a lack of relation between these variables.

With N = 7 this r will occur about 90% of the time by chance alone

SPSS correlation results		Fear of statistics
Social Isolation	Pearson Correlation	.058
	Sig. (2-tailed)	.902
	N	7



Dr. David McKirnan, davidmck@uic.edu

Critical values of Pearson Correlation (r)

N- 2 (N= number of pairs)	Level of significance for two-tailed test			
	p<.10	p<.05	p<.02	p<.01
1	.988	.997	.9995	.9999
2	.900	.950	.980	.990
3	.805	.878	.934	.959
4			.882	.917
5			.833	.874
10	.497	.576	.658	.708
15	.412	.482	.558	.606
25	.323			.487
30	.296	.349	.409	.449
60	.211	.250	.295	.325
70	.195	.232	.274	.302
80	.183	.217		
90	.173	.205		
100	.164	.195	.230	.254

Critical values of r

N = 27, alpha = .01

> 90 participants, alpha = .05

Like t, we test the statistical significance of a correlation by comparing it to a critical value of r.

We use N - 2…

…and the significance level (alpha)

… to find our critical value

Dr. David McKirnan, davidmck@uic.edu

Critical values of Pearson Correlation (r)

N- 2 (N= number of pairs)	Level of significance for two-tailed test			
	p<.10	p<.05	p<.02	p<.01
1	.988	.997	.9995	.9999
2	.900	.950	.980	.990
3	.805	.878	.934	.959
4	.729	.811	.882	.917
5	.669	.754	.833	.874
10	.497	.576	.658	.708
15	.412	.482	.558	.606
100	.164	.195	.230	.254

- ✓ The r we observed in *our* sample = .058.
- ✓ That is far less than the critical value of .75,
- ✓ Thus, we assume it is not significantly different from 0…
- ✓ …and we accept the null hypothesis that this occurred by chance alone.

Critical values of r:

We have n = 7;

N pairs-2 = 5 df

We test the effect at p<.05

The critical value of r with n = 7 (df = 5) is .75.

Correlation example 2

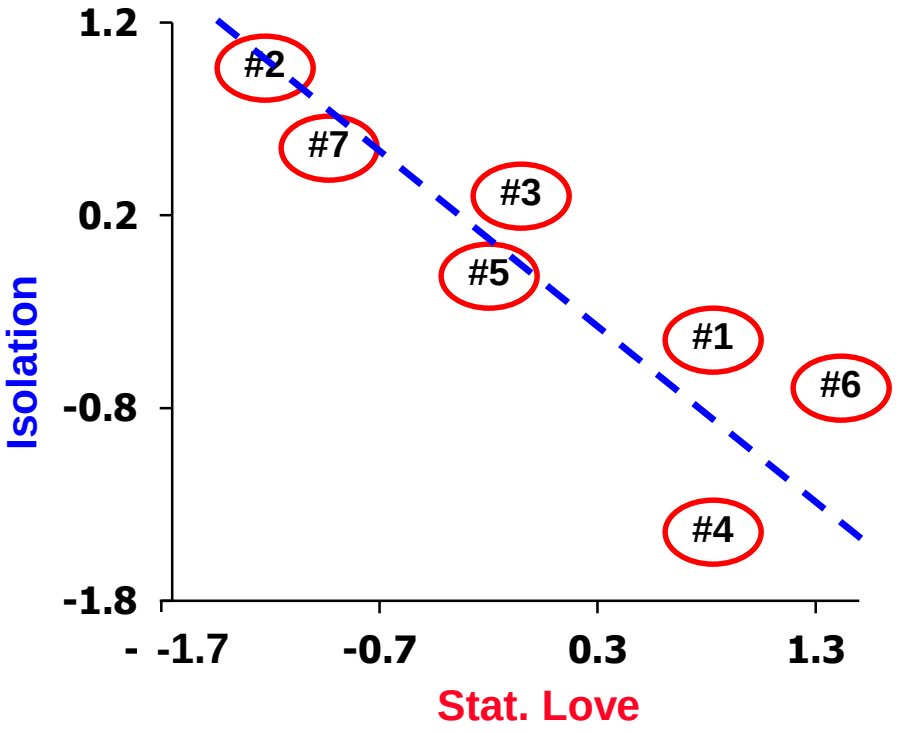
These data show a stronger relation between Love & Performance.

As with the first data set, we transform the raw scores to Zs

Example data set 2; Raw data		
Participant	Stat. Love ('X')	Isolation ('Y')
1	.2478	-.6564
2	1.1117	-1.3559
3	.2807	-.1406
4	-1.7773	.6563
5	.6229	-1.2559
6	-.5756	1.6388
7	.6229	-.8345

The scatter plot now shows strong linear relation;

a nearly ideal regression line.



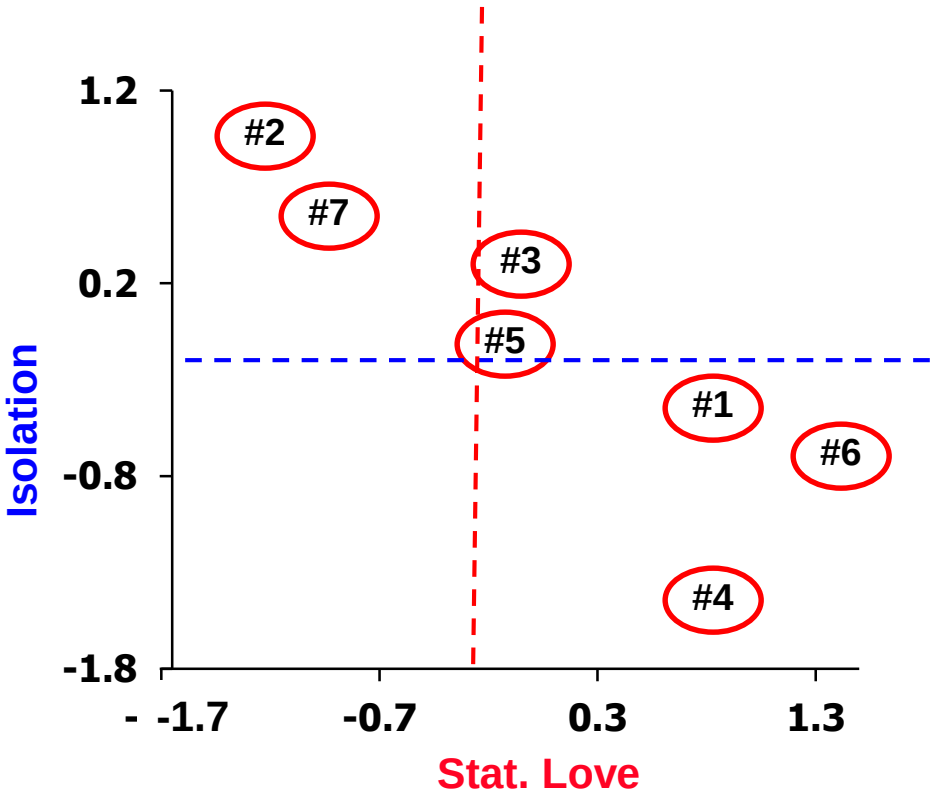
Dr. David McKirnan, davidmck@uic.edu

Correlation example 2

Participants who are above the Mean on Stat. Love tend to be below the M on Isolation.

Isolation decreasing as love of Statistics increases makes this a negative correlation.

The strength of the relation between Stat. Love and Isolation will be reflected in their mutual Z scores.



Calculating r

1. Calculate the Mean.
2. Calculate deviations, square them, and sum.
4. …and the Standard Deviation:

Participant	Social isolation ("Y")		Calculate the Sum of Squared deviations	Compute the standard deviation
	X	M		
1	4	4.57		2.05
2	7	4.57		2.05
3	5	4.57		2.05
4	1	4.57		2.05
5	6	4.57	= 2.04	2.05
6	3	4.57	= 2.46	2.05
7	6	4.57	= 2.04	2.05
(N = 7)	Σ = 32		**Σ (M – X)² = 25.35**	$S = \sqrt{\frac{\sum (X - M)^2}{df}}$

$M = \sum X / n = 32/7 = 4.57$

$= \sqrt{\frac{25.35}{6}} = \sqrt{4.22} = 2.05$

$(4 - 4.57)^2 =$

$(7 - 4.57)^2 =$

$(5 - 4.57)^2 = .$

Dr. David McKirnan, davidmck@uic.edu

3. Taking each participant's Score, the Mean, and the Standard Deviation compute Z for each participant.

Participant	Social isolation (variable "Y")		Standard Deviation (S)	Compute Z scores [Z = *X–M / S*]
	Score	M		
1	4	4.57	2.05	$Z = \frac{4 - 4.57}{2.05} = .278$
2	7	4.57	2.05	$Z = \frac{7 - 4.57}{2.05} = 1.87$
3	5	4.57	2.05	-.207
4	1	4.57	2.05	1.72
5	6	4.57	2.05	- .69
6	3	4.57	2.05	.76
7	6	4.57	2.05	-.69

Calculating r from Z scores

4. Enter the scores for each participant on the 2nd Variable.
…and compute the Z scores

5. Multiply the Zs from each variable for each participant.

sum, divide by n: $r = \frac{\Sigma (Z_Y * Z_X)}{n} = \frac{-5.69}{7} = -.81$

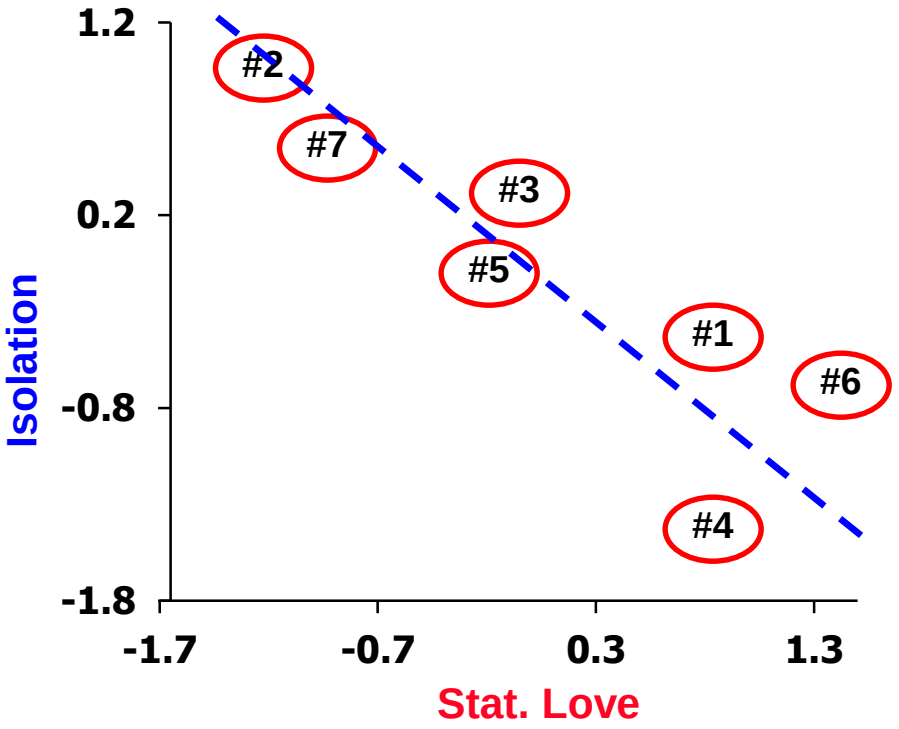
Participant	Social isolation ("Y")		Love of statistics ("X")		$Z_Y * Z_X$
	Score	Z score	Score	Z score	
1	4	.278	5	-.64	278 * -.64 = - .18
2	1	1.17	2	-1.59	1.17 * -1.59 = -1.87
3	3	.207	4	-.106	= - .02
4	7	-1.73	5	.637	= -1.09
5	2	.69	2	-1.59	= - .90
6	5	-.76	6	1.381	= -1.05
7	2	.69	3	-.845	= - .58
(n = 7)					$\Sigma (Z_Y * Z_X) = -5.69$

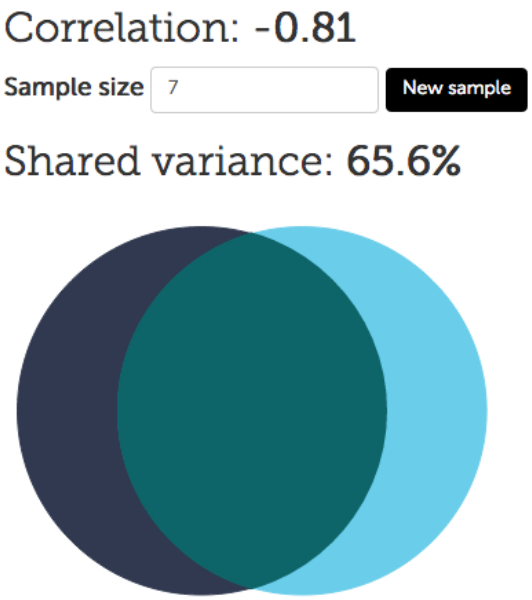
Dr. David McKirnan, davidmck@uic.edu

Correlation example 2

Where r = -.81 the scatter plot shows a tight distribution…

Scores on isolation go up / down as scores on Statistical Love do.





With r = -.81, r^2 = .656.

StatLove and isolation share a lot of variance.

Dr. David McKirnan, davidmck@uic.edu

Critical values of Pearson Correlation (r)

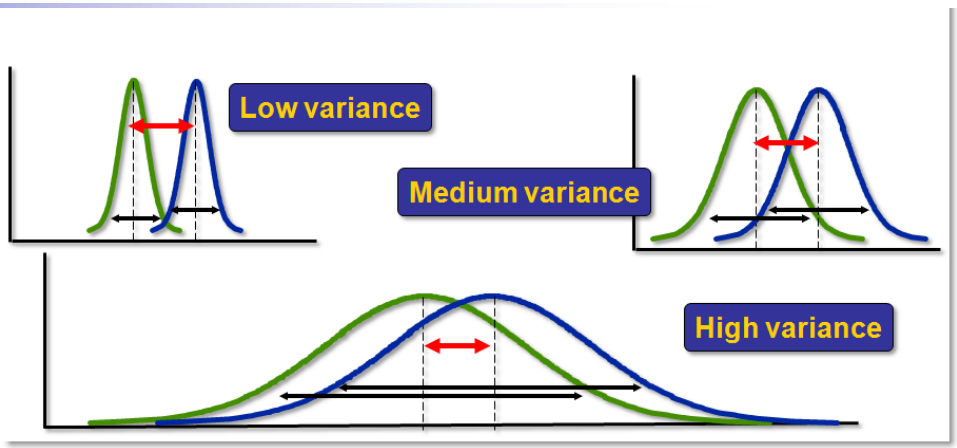
n- 2 (n = number of participants)	Level of significance for two-tailed test			
	p<.10	p<.05	p<.02	p<.01
1	.988	.997	.9995	.9999
2	.900	.950	.980	.990
3	.805	.878	.934	.959
4	.729	.811	.882	.917
5	.669	.754	.833	.874
10	.497	.576	.658	.708
15	.412	.482	.558	.606
2				
3				
6				
7				
8				
9				
1(				

Testing r against the Critical value:

- ✓ r in *this* sample = - 81.
- ✓ Since it is < .75, the correlation is statistically significant at p<.05.
- ✓ We reject the null hypothesis that this strong a relation occurred by chance alone…
- ✓ …and conclude that loving statistics does, in fact, make you a better & happier person.

Summary: Statistical tests

t-test: Compare one group to another



Experimental v. control (Experiment)

Men v. women, etc. (Measurement)

- Calculate M for each group, compare them to determine how much variance is due to differences between groups.
- Calculate standard error to determine how much variance is due to individual differences within each group.
- Calculate the critical ratio (t):

$$\frac{\text{Difference between groups}}{\text{standard error of M}}$$

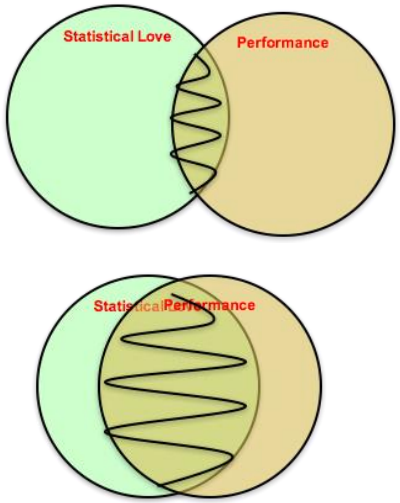
Dr. David McKirnan, davidmck@uic.edu

Pearson Correlation (r): how similar is the variance between two variables within a sample.

- Are people above or below the mean on one variable similarly above or below M on the second variable?

r = the product of each participant's Z score for two variables / n:

$$r = \frac{\sum Z_X * Z_Y}{n}$$



r^2 measures how much variance is shared between two variables.

✓ Statistical notation so far…

X	Score on one variable for one participant	
n	Number of scores in the sample	
Σ	Sum of a set of scores	
M or	Mean; sum of scores divided by n of scores:	$\sum x / n$
X - M	deviation of one score from the mean	
$(X - M)^2$	squared deviation of score from mean	
SS	sum of squared deviations from the mean:	$\Sigma (X - M)^2$
df	degrees of freedom; # of scores that are free to vary:	n - 1
S^2	Variance: sum of squared deviations from *M* divided by degrees of freedom:	$\sum (X - M)^2 / n - 1$
S	Standard Deviation, square root of the variance.	
Z	# standard deviation units from M:	X – M / S
t	Between group variance / within group variance.	$t = \frac{(M_{group1} - M_{group2}) - 0}{\sqrt{\frac{\sum (X - M)^2_{grp1} / n - 1_{grp1}}{n_{grp1}} + \frac{\sum (X - M)^2_{grp2} / n - 1_{grp2}}{n_{grp2}}}}$	

r **Similarity in variance between 2 variables**

r^2 **Shared variance between 2 variables**

$$r = \frac{\sum Z_X * Z_Y}{n}$$

Type I v. Type II errors in research decisions

- ✓ **Statistical decisions are probability judgements:**
  - ➢ **Large critical ratios have a very low probability of occurring by chance alone…**
  - ➢ **…but it is still always possible that what we are seeing is by chance.**
- ✓ **That is, our conclusions may always be by error.**
- ✓ **How do we think about error in research?**

		[effect due to chance alone]	[real experimental effect]
Decision	**Accept Ho**		
	Reject Ho		

- ✓ Do we *accept* or *reject* the null hypothesis (Ho)?
- ✓ If the results are **chance alone** – there is no real effect – we are correct to **accept Ho**.
- ✓ If the IV had a **real effect** on the DV, we correctly **reject the Ho**.

		"Reality"	
		✓ Ho true Effect = **chance** alone	✓ Ho false Effect is **Real**
Decision	Accept Ho	**Correct decision**	
	Reject Ho		**Correct decision**

- ✓ Do we *accept* or *reject* the null hypothesis (Ho)?
- ✓ ***Rejecting* Ho** (deciding there is a real effect) **when results are due to chance = Type I error.**
- ✓ ***Accepting* Ho when results are due to a real effect = Type II error.**

		"Reality"	
		Ho true Effect = **chance** alone	**Ho false** Effect is Real
Decision	**Accept Ho**	Correct decision	**Type II error**
	Reject Ho	**Type I error**	Correct decision

Dr. David McKirnan, davidmck@uic.edu

- ✓ Do we *accept* or *reject* the null hypothesis (Ho)?
- ✓ **Type I error: controlled by alpha (α).**
- ✓ **α = ,05 means we make Type I errors 5% of the time.**

		"Reality"	
		Ho true [effect due to chance alone]	**Ho false** [real experimental effect]
Decision	**Accept Ho**	**Correct decision 1 - α**	
	Reject Ho	**Type I error α**	

We typically set this at p<.05

Alpha (e.g., α = p<.05) **= probability of Type I error** (5%)**.**

α = p<.01 is more conservative: 1% Type I error rate.

Dr. David McKirnan, davidmck@uic.edu

Type I error; Reject the null hypothesis [Ho] when it is actually true:

- Accept as 'real' an effect that is due to chance only
- Type I error rate is determined by Alpha (.10 → .05 → .01 → .001…)
- As we choose a less stringent alpha level, critical values get lower.
- This makes it easier to commit a Type I error.
- Type I is the "worst" error; our statistical conventions are designed to prevent these.

	Alpha Levels				
df	0.10	0.05	0.02	0.01	0.001
8	1.860	2.306	2.896	3.355	5.041
9	1.833	2.262	2.821	3.250	4.781
10	1.812	2.228	2.764	3.169	4.587
11	1.796	2.201	2.718	3.106	4.437
12	1.782	2.179	2.681	3.055	4.318
13	1.771	2.160	2.650	3.012	4.221
14	1.761	2.145	2.624	2.977	4.140
15	1.753	2.131	2.602	2.947	4.073
20	1.725	2.086	2.528	2.845	3.850
25	1.708	2.060	2.485	2.787	3.725
30	1.697	2.042	2.457	2.750	3.646
40	1.684	2.021	2.423	2.704	3.551
60	1.671	2.000	2.390	2.660	3.460
120	1.658	1.980	2.358	2.617	3.373
∞	1.645	1.960	2.326	2.576	3.291

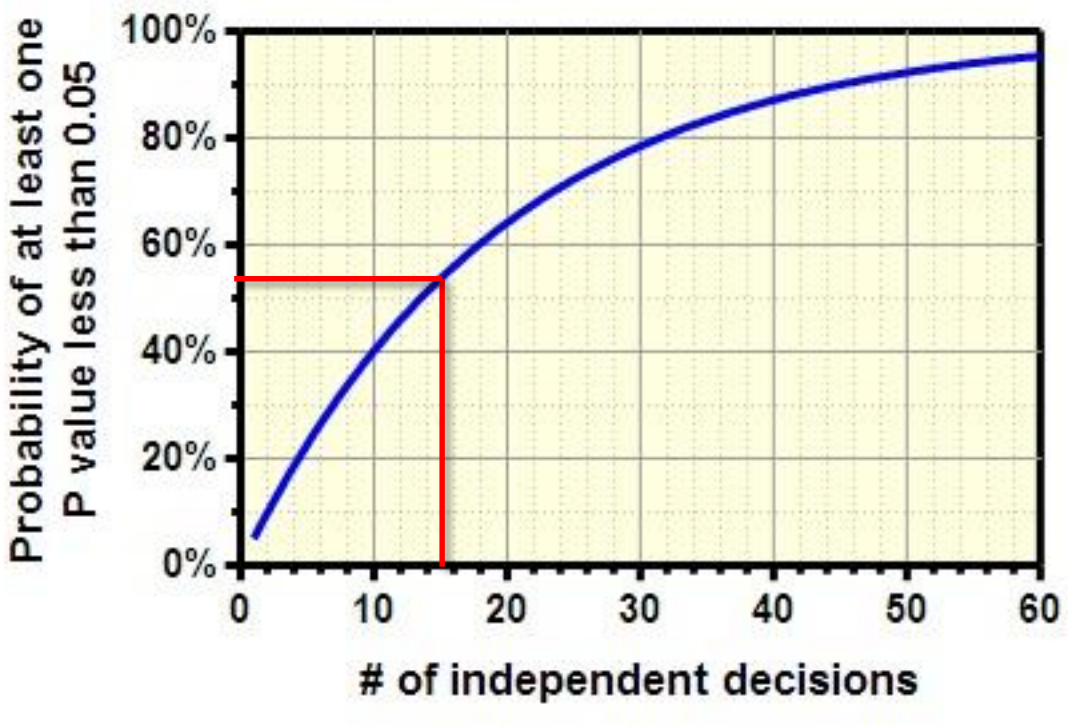
Type I error; Reject null hypothesis [Ho] when it is actually true:

- **Type I errors also increase as we run more tests.**
  - **With a single statistical test** (t-test, Z, r…) **at p<.05:**
    - Probability (P) of Type I error = alpha (α) = 5%.
    - **V**(Probability of Correctly accept Ho) = (1 - α) = 95%.
  - **With 3 statistical tests:**
    - P(≥ 1 significant result by chance) = $(1 - V)^{\text{number of tests}}$
    - $P = 1 - (1 - .05)^3 = 1 - (.95)^3 = 1 - .86 \approx .14$
    - So, if we run 3 different t-tests at p<.05, the odds of any one appearing to be significant in junk data is 14%.

Type I error; Reject null hypothesis [Ho] when it is actually true:

- **With 13 statistical tests:**
  - $P = 1 - (1 - .05)^{13} = 1 - .95^{13} = 1 - .51 \approx .50\%$
  - With 13 tests in random data the odds of any one being "statistically significant" [a type I error) is about 50%.
  - Half the time we will mistake junk variance for a true effect.

As we run more statistical tests, Type I error increases substantially



- ✓ Do we *accept* or *reject* the null hypothesis (Ho)?
- ✓ **Type II errors**

		"Reality"	
		Ho true [chance]	**Ho false** [real experimental effect]
Decision	**Accept Ho**	decision 1 - α	**Type II error β**
	Reject Ho	Type I error α	**Correct decision 1 - β**

Determined by the Power of our experiment.

Alpha (α, typically p<.05) = probability of Type I error.

β = probability of Type II error.

Dr. David McKirnan, davidmck@uic.edu

Type II error; Accept Ho when it is actually false;

- Type II rate most strongly affected by statistical power.

Central Limit Theorem:

Smaller samples ⇨ Assume more variance ⇨ More conservative critical value

- Smaller samples have more stringent critical values…
- So they have less 'power' to detect a significant effect.
- In small "underpowered" samples it is difficult for even strong effects to come out statistically significant.

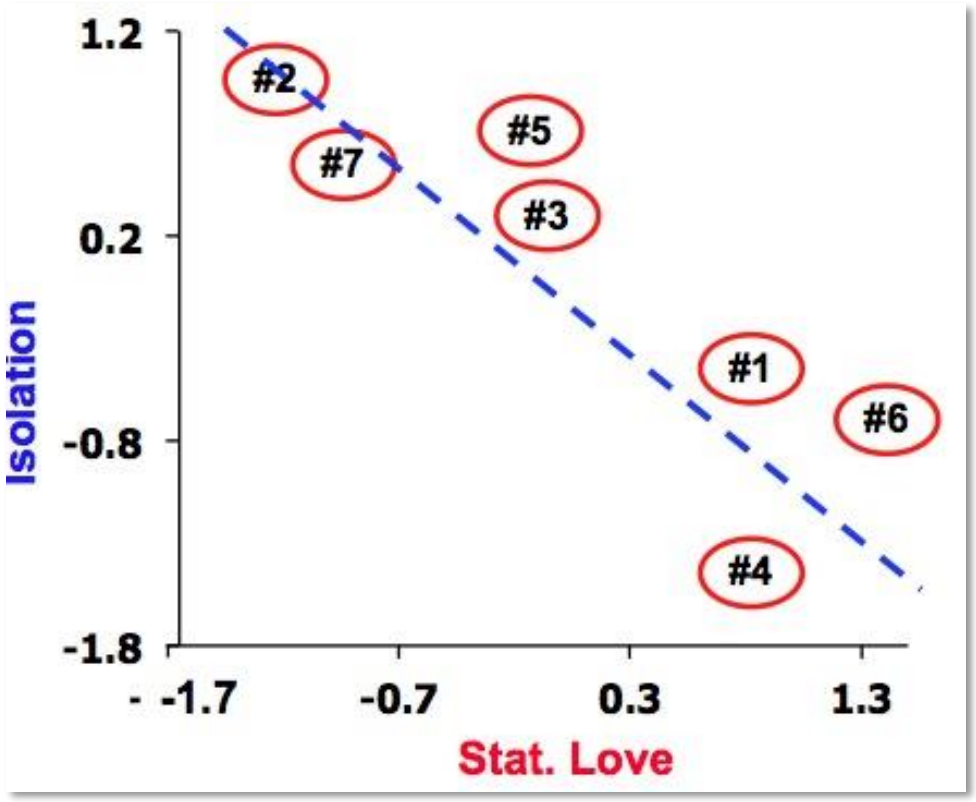
	Alpha Levels				
df	0.10	0.05	0.02	0.01	0.001
8	1.860	2.306	2.896	3.355	5.041
9	1.833	2.262	2.821	3.250	4.781
10	1.812	2.228	2.764	3.169	4.587
11	1.796	2.201	2.718	3.106	4.437
12	1.782	2.179	2.681	3.055	4.318
13	1.771	2.160	2.650	3.012	4.221
14	1.761	2.145	2.624	2.977	4.140
15	1.753	2.131	2.602	2.947	4.073
20	1.725	2.086	2.528	2.845	3.850
25	1.708	2.060	2.485	2.787	3.725
30	1.697	2.042	2.457	2.750	3.646
40	1.684	2.021	2.423	2.704	3.551
60	1.671	2.000	2.390	2.660	3.460
120	1.658	1.980	2.358	2.617	3.373
∞	1.645	1.960	2.326	2.576	3.291

- ✓ Do we *accept* or *reject* the null hypothesis (Ho)?
- ✓ **We use experimental design…**
  - **Avoiding confounds,**
  - **Standardizing and controlling procedures,**
  - **Haivng a clear sampling frame…**
- **… & statistical conventions to avoid these errors.**

		"Reality"	
		Ho true Effect = **chance** alone	**Ho false** Effect is Real
Decision	**Accept Ho**	Correct decision	**Type II error**
	Reject Ho	**Type I error**	Correct decision

Illustrations of Type I & Type II errors from the Statistical Love – Social isolation data.

Imagine r for these data = -.69



Type I error:

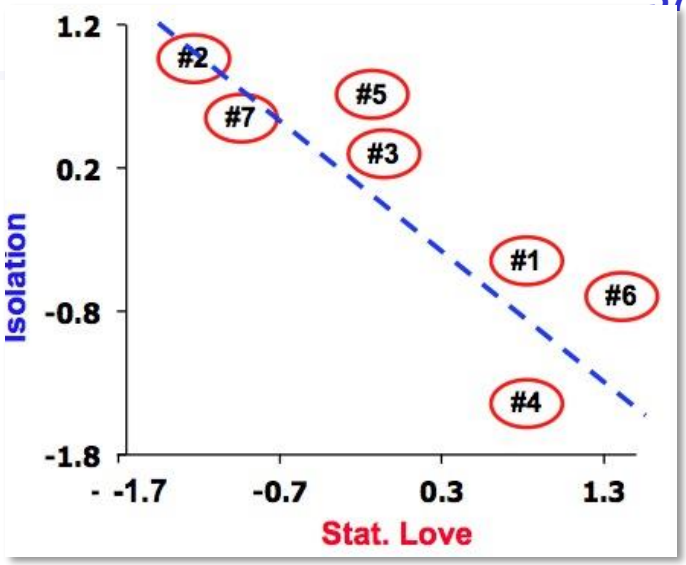
Our r value (.69) did not exceed the critical value for Alpha = p<.05, with 5 df.

However it did exceed the critical value for p<.10.

We took this as supporting a "trend" in the data.

By adopting the lower Alpha value, we may be committing a Type I error;

We may be rejecting the null hypothesis where there really is not an effect.



n- 2 (n = number of participants)	Level of significance for two-taile		
	p<.10	p<.05	p<.02
1	.988	.997	.9995
2	.900	.950	.980
3	.805	.878	.934
4	.729	.811	.882
5	.669	.754	.833
10	.497	.576	.658
15	.412	.482	.558
25	.323	.381	.445
30	.296	.349	.409

Type II error:

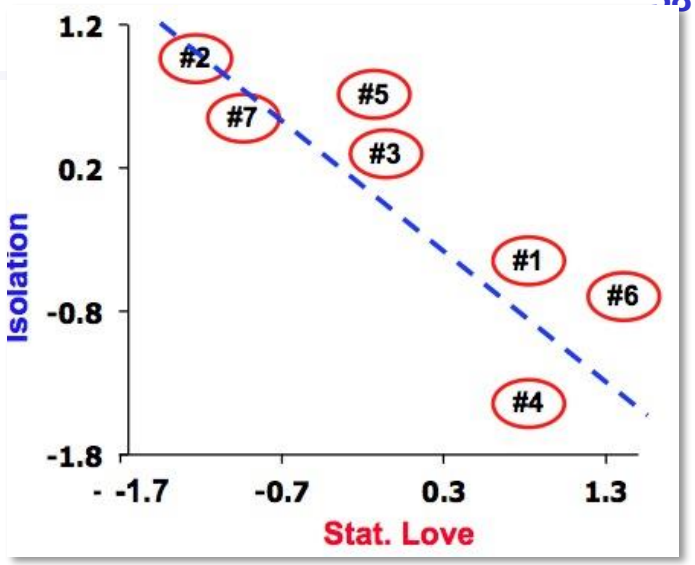
From the graph Statistical Love is clearly related to lessened Social Isolation.

However, with only 5 degrees of freedom we have a very stringent critical value for r.

At even 10 df the critical value goes down substantially.

Our study is so 'underpowered' that it is almost impossible for us to find an effect (reject the null hypothesis)**.**

We need more statistical power to have a fair test of the hypothesis.



n- 2 (n = number of participants)	Level of significance for two-taile		
	p<.10	p<.05	p<.02
1	.988	.997	.9995
2	.900	.950	.980
3	.805	.878	.934
4	.729	.811	.882
5	.669	.754	.833
10	.497	.576	.658
15	.412	.482	.558
25	.323	.381	.445
30	.296	.349	.409

Type I error; Reject the null hypothesis [Ho] when it is actually true:

- Accept as 'real' an effect that is due to chance only
- Type I error rate is determined by Alpha (.10 → .05 → .01 → .001…)
- As we choose a less stringent alpha level, critical values get lower.
- This makes it easier to commit a Type I error.
- Type I is the "worst" error; our statistical conventions are designed to prevent these.

	Alpha Levels				
df	0.10	0.05	0.02	0.01	0.001
8	1.860	2.306	2.896	3.355	5.041
9	1.833	2.262	2.821	3.250	4.781
10	1.812	2.228	2.764	3.169	4.587
11	1.796	2.201	2.718	3.106	4.437
12	1.782	2.179	2.681	3.055	4.318
13	1.771	2.160	2.650	3.012	4.221
14	1.761	2.145	2.624	2.977	4.140
15	1.753	2.131	2.602	2.947	4.073
20	1.725	2.086	2.528	2.845	3.850
25	1.708	2.060	2.485	2.787	3.725
30	1.697	2.042	2.457	2.750	3.646
40	1.684	2.021	2.423	2.704	3.551
60	1.671	2.000	2.390	2.660	3.460
120	1.658	1.980	2.358	2.617	3.373
∞	1.645	1.960	2.326	2.576	3.291

Type II error; Accept Ho when it is actually false;

- Assume as chance an effect that is actually real.
- Type II rate most strongly affected by statistical power.

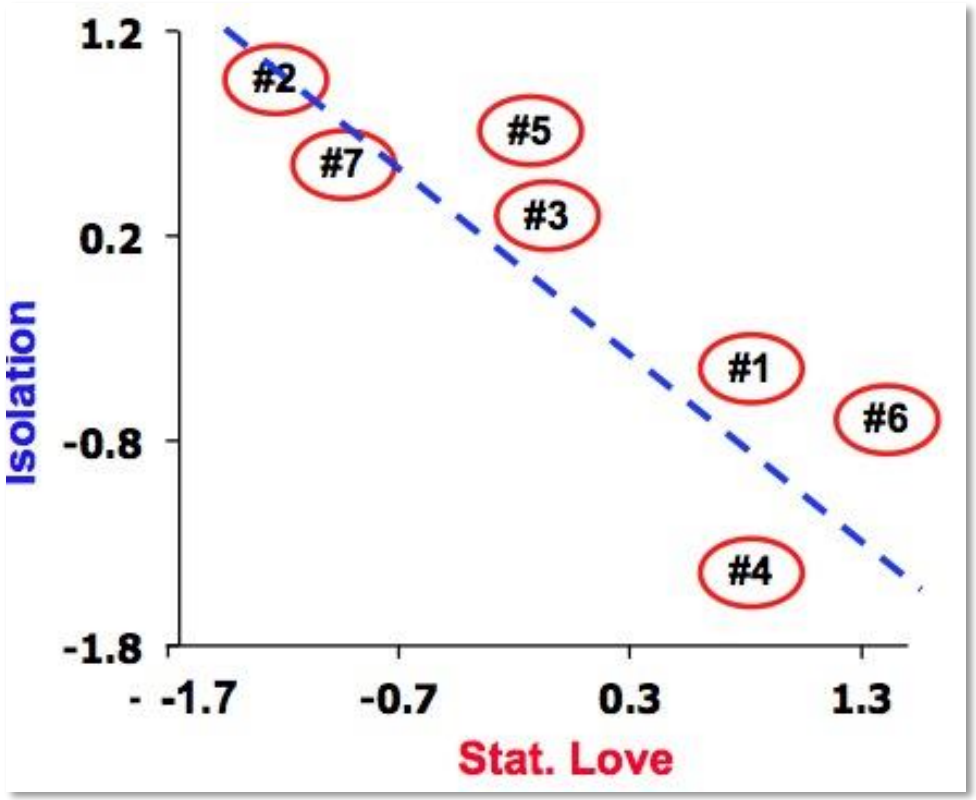
Central Limit Theorem:

Smaller samples ⇨ Assume more variance ⇨ More conservative critical value

- Smaller samples have more stringent critical values…
- So they have less 'power' to detect a significant effect.
- In small "underpowered" samples it is difficult for even strong effects to come out statistically significant.

	Alpha Levels				
df	0.10	0.05	0.02	0.01	0.001
8	1.860	2.306	2.896	3.355	5.041
9	1.833	2.262	2.821	3.250	4.781
10	1.812	2.228	2.764	3.169	4.587
11	1.796	2.201	2.718	3.106	4.437
12	1.782	2.179	2.681	3.055	4.318
13	1.771	2.160	2.650	3.012	4.221
14	1.761	2.145	2.624	2.977	4.140
15	1.753	2.131	2.602	2.947	4.073
20	1.725	2.086	2.528	2.845	3.850
25	1.708	2.060	2.485	2.787	3.725
30	1.697	2.042	2.457	2.750	3.646
40	1.684	2.021	2.423	2.704	3.551
60	1.671	2.000	2.390	2.660	3.460
120	1.658	1.980	2.358	2.617	3.373
∞	1.645	1.960	2.326	2.576	3.291

Illustrations of Type I & Type II errors from the Statistical Love – Social isolation data



Type I error:

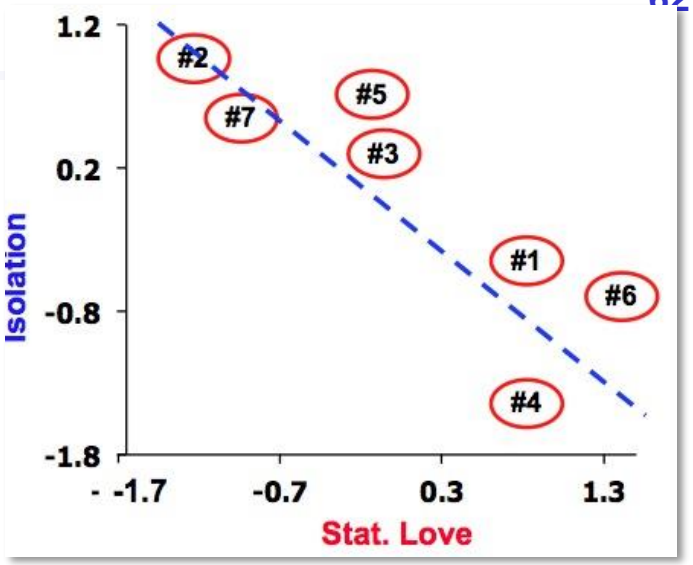
Our r value (.69) did not exceed the critical value for Alpha = p<.05, with 5 df.

However it did exceed the critical value for p<.10.

We took this as supporting a "trend" in the data.

By adopting the lower Alpha value, we may be committing a Type I error;

We may be rejecting the null hypothesis where there really is not an effect.



n- 2 (n = number of participants)	Level of significance for two-taile		
	p<.10	p<.05	p<.02
1	.988	.997	.9995
2	.900	.950	.980
3	.805	.878	.934
4	.729	.811	.882
5	.669	.754	.833
10	.497	.576	.658
15	.412	.482	.558
25	.323	.381	.445
30	.296	.349	.409

Type II error:

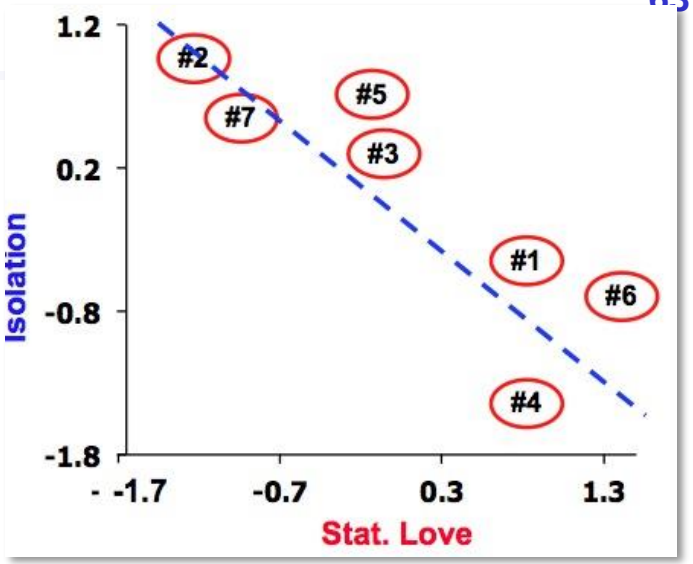
From the graph Statistical Love is clearly related to lessened Social Isolation.

However, with only 5 degrees of freedom we have a very stringent critical value for r.

At even 10 df the critical value goes down substantially.

Our study is so 'underpowered' that it is almost impossible for us to find an effect (reject the null hypothesis)**.**

We need more statistical power to have a fair test of the hypothesis.



n- 2 (n = number of participants)	Level of significance for two-taile		
	p<.10	p<.05	p<.02
1	.988	.997	.9995
2	.900	.950	.980
3	.805	.878	.934
4	.729	.811	.882
5	.669	.754	.833
10	.497	.576	.658
15	.412	.482	.558
25	.323	.381	.445
30	.296	.349	.409

- ✓ **Our research mission is to develop explanations – theories – of the natural world.**
- ✓ **Many of the basic processes we are interested in are hypothetical constructs, that we cannot simply 'see' directly.**
- ✓ **We develop hypotheses and collect data, then use statistics and inferential logic to help us minimize the Type I and Type II errors that plague human judgment.**

Dr. David McKirnan, davidmck@uic.edu

Core assumptions of inferential statistics

- **We impute a Sampling Distribution based on the variance, and the Degrees of Freedom** [*df*] **of our sample.**
- **This is a distribution of possible results – that is, potential critical ratios** (t scores, r…) **– based on our data.**
- **We compare our results to the sampling distribution to estimate the probability that they are due to a "real" experimental effect rather than chance or error.**
- **The Central Limit Theorem tells us that data from smaller samples** (fewer *df*) **will have more error variance.**
- **So, with fewer *df* we estimate a more conservative sampling distribution; the critical value we test our t or r against gets higher.**

Type I v. Type II errors

Decreased by more statistical power (↑ participants)

		"Reality"	
		Ho true [effect due to chance alone]	**Ho false** [real experimental effect]
Decision	**Accept Ho**	**Correct decision**	**Type II error**
	Reject Ho	**Type I error**	**Correct decision**

Decreased by more conservative Alpha ($p < .05 \rightarrow .01 \rightarrow .001$)

Inferential statistics: summary, **Key terms**

- **Plato's cave and the estimation of "reality"**
- **Inferences about our observations:**
  - **Deductive v. Inductive link of theory / hypothetical constructs & data**
  - **Generalizing results beyond the experiment**
- **Critical ratio (you *will* be asked to produce and describe this).**
- **Variance, variability in different distributions**
- **Degrees of Freedom [df]**
- **t-test, between versus within –group variance**
- **Sampling distribution, M of the sampling distribution**
- **Alpha (α), critical value**
- **t table, general logic of calculating a t-test**
- **"Shared variance", positive / negative correlation**
- **General logic of calculating a correlation (mutual Z scores).**
- **Null hypothesis, Type I & Type II errors.**